Introduction to FUNCTIONAL

ANGUS E. TAYLOR

Professor of Mathematics
University of California
Los Angeles

ANALYSIS

New York · John Wiley & Sons, Inc.
London · Chapman & Hall, Ltd.

Copyright © 1958 by John Wiley & Sons, Inc.

All Rights Reserved. This book or any part thereof must not be reproduced in any form without the written permission of the publisher.

Library of Congress Catalog Card Number: 58–12704

Printed in the United States of America

To GORDON,

KENNETH, and KITTY

PREFACE

This book has been written as a part of my program of teaching at the graduate level. The primary aim of the book is to assist graduate students in learning fundamental ideas and theorems about linear spaces and linear operators and to lead them to an appreciation of the unifying power of the abstract-linear-space point of view in surveying the problems of algebra, classical analysis, the theory of integration, and differential and integral equations. While the book is principally addressed to graduate students, it is also intended to be useful to mathematicians, both pure and applied, who have need of a simple and direct presentation of the fundamentals of the theory of linear spaces and linear operators.

The central theme of the book is the theory of normed linear spaces and of linear operators which map one such space into another. A normed linear space which is complete is called a Banach space; a number of the most important results in linear-space theory depend on completeness, and so we have a Banach space context for these results. However, the hypothesis of completeness is not invoked except as it is needed for effective results. The more specialized theory of inner-product spaces (which, when complete and infinite dimensional are called Hilbert spaces) is carried along within the normed-linear-space context, leaning upon the general theory of normed linear spaces, but with its own characteristic special development where such development can achieve results of a distinctive character.

Although the emphasis is mainly on normed linear spaces, the more general ideas of topological linear spaces are developed to some extent, since an understanding of these is relevant at many places in the theory of normed linear spaces, notably in connection with the weak topologies of such spaces and their conjugate spaces. Part of Chapter 3 provides a brief introduction to the study of general topological linear spaces, both without and with the assumption of local convexity. The geometry of convex sets and linear varieties (especially hyperplanes) is all presented in the context of topological linear spaces.

There are many illustrations and applications of the abstract concepts and methods. The numerous problems serve both to illustrate and to extend the theory, as well as to train the student by having him fill in details omitted in the text. For many of the illustrations and problems the sequence spaces l^p have been chosen, in order to minimize the technical analytical difficulties. There are applications to differential equations, to integral equations, and to the classical theory of analytic functions.

Chapter 5 stresses the importance of complex contour integration and the calculus of residues in the spectral theory of linear operators. The methods apply to all closed linear operators, bounded or not. This chapter also contains the famous Riesz theory of compact (completely continuous) operators, as extended and perfected by later research workers, and its application to obtain the classical "determinant-free" theorems for Fredholm integral equations of the second kind.

Chapter 6 presents the standard elementary theory of self-adjoint, normal, and unitary operators in Hilbert space. There is also the discussion, very important for applications to integral and differential equations, of the theory of compact symmetric operators and symmetric operators with compact resolvent. This discussion does not require the completeness of the inner-product space under consideration. The spectral analysis of self-adjoint operators is performed with the aid of the Riesz representation theorem for linear functionals on a space of continuous functions. The treatment is deliberately kept as close to classical analysis as possible within this framework. After the student is acquainted with the situation from this point of view, he may with great profit go on to a study of B*-algebras, and learn to view the spectral theorem for a normal or self-adjoint operator as a consequence of the Gelfand-Neumark theorem for commutative B*-algebras. For an understanding of these developments Chapter 7 is an essential preparation. But there is no room in this book for a treatment of Banach algebras.

The book is an introduction, not a treatise. It is meant to open doors for the student and to give him understanding and preparation which will help him to push on, if he wishes, to the new frontiers of modern mathematics, carrying with him a clearer realization of the structure of classical mathematics.

ANGUS E. TAYLOR

Los Angeles
August 1957

ACKNOWLEDGMENTS

The first draft of this book was written while I was on sabbatical leave from the University of California, Los Angeles. During part of that time I worked in Geneva, where I was accorded the courtesy of the libraries of the University and of the Mathematical Institute. Then for six months in 1955 I was privileged to be a Fulbright Research Scholar visiting at the Johannes Gutenberg University in Mainz, Germany. I am deeply grateful to Professor Gottfried Köthe, who was at that time Rektor of the University and who graciously furnished me his own office in the Mathematical Institute.

I wish to thank the University of California for a small grant of funds for assistance in the preparation of the manuscript. I also thank Ruthanne Clark for very expert typing and handwork on the formulas.

The book has taken shape in my mind gradually since 1946. I owe a great deal to successive groups of graduate students to whom I have lectured and who have participated in seminars. For various improvements in style and presentation and for the detection of errors of omission and commission, I am especially indebted to my 1956–1957 class, which read the manuscript in its nearly final form.

For helpful conversations and comments I am especially indebted to Professors Richard Arens and Magnus Hestenes. Lastly, I take pleasure in thanking Professor Nelson Dunford for providing me with a manuscript copy of the book by himself and J. T. Schwartz. This voluminous and encyclopedic manuscript, available to me after my own first draft was written, made it easier for me to resign myself to the omission of topics which I had originally hoped to be able to include in my own book.

A. E. T.

CONTENTS

INTRODUCTION 1

SETS 1
FUNCTIONS 4
INVERSE FUNCTIONS 5
USAGE RELATING TO REAL NUMBERS 5
INEQUALITIES 5
THE KRONECKER DELTA 6

CHAPTER 1 THE ABSTRACT APPROACH TO LINEAR PROBLEMS 7

1.0 INTRODUCTORY STATEMENT 7
1.1 ABSTRACT LINEAR SPACES 8
1.2 EXAMPLES OF LINEAR SPACES 13
1.3 LINEAR OPERATORS 17
1.4 LINEAR OPERATORS IN FINITE DIMENSIONAL SPACES 22
1.5 OTHER EXAMPLES OF LINEAR OPERATORS 26
1.6 LINEAR FUNCTIONALS 33
1.61 LINEAR FUNCTIONALS IN FINITE DIMENSIONAL SPACES 37
1.7 ZORN'S LEMMA 39
1.71 EXTENSION THEOREMS FOR LINEAR OPERATORS 40
1.72 HAMEL BASES 44
1.8 THE TRANSPOSE OF A LINEAR OPERATOR 46
1.9 ANNIHILATORS 48

1.91 Ranges and null manifolds 52
1.92 Conclusions 54

CHAPTER 2 TOPOLOGIES 56

2.0 Scope of the chapter 56
2.1 Topological spaces 56
2.11 Relative topologies 60
2.12 Continuous functions 61
2.2 Compact sets 62
2.21 Category. Separability 63
2.3 Separation axioms. Hausdorff spaces 64
2.31 Locally compact spaces 67
2.4 Metric spaces 68
2.41 Completeness 74
2.5 Product spaces 79

CHAPTER 3 TOPOLOGICAL LINEAR SPACES 81

3.0 Introduction 81
3.1 Normed linear spaces 83
3.11 Examples of normed linear spaces 87
3.12 Finite-dimensional normed linear spaces 94
3.13 Banach spaces 98
3.14 Quotient spaces 104
3.2 Inner-product spaces 106
3.21 Hilbert spaces 118
3.22 The completeness of certain orthonormal sets 121
3.3 Topological linear spaces 123
3.4 Convex sets 130
3.41 Minkowski functionals 134
3.5 Linear varieties 137
3.6 Convex sets and hyperplanes 140
3.7 Seminorms 143
3.8 Locally convex spaces 145
3.81 Weak topologies for linear spaces. Duality 151
3.9 Metric linear spaces 154

CHAPTER 4 GENERAL THEOREMS ON LINEAR OPERATORS 160

4.0 Introduction 160
4.1 Spaces of linear operators 161
4.11 Integral equations of the second kind. The Neumann expansion 165

4.12 $\mathfrak{L}^2$ KERNELS 169
4.13 DIFFERENTIAL EQUATIONS AND INTEGRAL EQUATIONS 171
4.2 CLOSED LINEAR OPERATORS 175
4.3 THE NORMED CONJUGATE OF A NORMED LINEAR SPACE 185
4.31 THE SECOND NORMED CONJUGATE SPACE 191
4.32 SOME REPRESENTATIONS OF LINEAR FUNCTIONALS 193
4.4 THE PRINCIPLE OF UNIFORM BOUNDEDNESS 201
4.41 WEAK CONVERGENCE 208
4.42 AN APPLICATION OF VECTOR-VALUED ANALYTIC FUNCTIONS 210
4.5 THE CONJUGATE OF A BOUNDED LINEAR OPERATOR 213
4.51 SOME REPRESENTATIONS OF BOUNDED LINEAR OPERATORS 215
4.52 THE M. RIESZ CONVEXITY THEOREM 221
4.6 ANNIHILATORS, RANGES, AND NULL MANIFOLDS 224
4.61 WEAK COMPACTNESS IN NORMED LINEAR SPACES 227
4.62 SATURATED SUBSPACES OF A CONJUGATE SPACE 231
4.7 THEOREMS ABOUT CONTINUOUS INVERSES 233
4.71 THE STATES OF AN OPERATOR AND ITS CONJUGATE 235
4.8 PROJECTIONS 240
4.81 CONTINUOUS LINEAR FUNCTIONALS ON A HILBERT SPACE 243
4.82 ORTHOGONAL COMPLEMENTS 245
4.83 DIRICHLET'S PRINCIPLE 246
4.9 ADJOINT OPERATORS 249

CHAPTER 5 SPECTRAL ANALYSIS OF LINEAR OPERATORS 253

5.0 INTRODUCTION 253
5.1 THE RESOLVENT OPERATOR 256
5.2 THE SPECTRUM OF A BOUNDED LINEAR OPERATOR 260
5.3 SUBDIVISIONS OF THE SPECTRUM 264
5.4 REDUCIBILITY 268
5.41 THE ASCENT AND DESCENT OF AN OPERATOR 271
5.5 COMPACT OPERATORS 274

5.6 THE OPERATIONAL CALCULUS 287
5.7 SPECTRAL SETS AND PROJECTIONS 298
5.71 THE SPECTRAL MAPPING THEOREM 302
5.8 ISOLATED POINTS OF THE SPECTRUM 305
5.9 OPERATORS WITH RATIONAL RESOLVENT 314

CHAPTER 6 SPECTRAL ANALYSIS IN HILBERT SPACE 321

6.0 INTRODUCTION 321
6.1 BILINEAR AND QUADRATIC FORMS 322
6.11 SYMMETRIC OPERATORS 324
6.12 A THEOREM OF SCHUR 327
6.2 NORMAL AND SELF-ADJOINT OPERATORS 329
6.3 ORTHOGONAL PROJECTIONS 333
6.4 COMPACT SYMMETRIC OPERATORS 334
6.41 SYMMETRIC OPERATORS WITH COMPACT RESOLVENT 342
6.5 THE SPECTRAL THEOREM FOR BOUNDED SELF-ADJOINT OPERATORS 345
6.6 UNITARY OPERATORS 356
6.7 UNBOUNDED SELF-ADJOINT OPERATORS 361

CHAPTER 7 INTEGRATION AND LINEAR FUNCTIONALS 365

7.0 EXPLANATORY REMARKS 365
7.1 THE SPACE $L(\mu)$ 366
7.2 SIGNED AND COMPLEX MEASURES 373
7.21 THE RADON-NIKODYM THEOREM 376
7.3 THE REAL SPACE $L^p(\mu)$ 376
7.4 CONTINUOUS LINEAR FUNCTIONALS ON L^p 381
7.41 COMPLEX L^p SPACES 385
7.5 MEASURES IN LOCALLY COMPACT HAUSDORFF SPACES 386
7.51 SIGNED AND COMPLEX BOREL MEASURES 388
7.6 VECTOR LATTICES 391
7.7 LINEAR FUNCTIONALS ON $C_\infty(T)$ 392
7.8 FINITELY ADDITIVE SET FUNCTIONS 401
7.9 LEBESGUE INTEGRATION WITH RESPECT TO A CHARGE 404

BIBLIOGRAPHY 407

LIST OF SPECIAL SYMBOLS 413

DISCUSSIONS RELATING TO PARTICULAR SPACES 415

INDEX 417

INTRODUCTION

It is the purpose of this introduction to explain certain notations and terminologies used throughout the book.

Sets

Let X be a given set. If x is an element of X this fact is expressed in symbols by writing

$$x \in X.$$

The negation of $x \in X$ is written in the form

$$x \notin X.$$

A subset of X is a set E such that if $x \in E$, then also $x \in X$. In particular, X is a subset of itself. A subset of X which is not all of X is called a *proper* subset of X. The statement "E is a subset of X" is expressed in symbols by writing

$$E \subset X, \qquad \text{or} \qquad X \supset E.$$

Two sets E, F are the same if and only if $E \subset F$ and $F \subset E$.

Among the subsets of a given set is the *empty set*, the set having no elements. We denote the empty set by $\emptyset$.

Let X be a set and let x be a variable element of X. For each x let $P(x)$ denote a proposition concerning x. Then we use the symbol

$$\{x:P(x)\}$$

to denote the set of all $x \in X$ for which $P(x)$ is true. This notation is extended in an obvious manner for more than one proposition. Thus, if $P_1(x)$, $P_2(x)$, $\cdots$ are propositions, $\{x:P_n(x),\ n=1, 2, \cdots\}$ denotes the set of all $x \in X$ such that $P_n(x)$ is true for each of the values $n = 1, 2, \cdots$.

Example 1. Let X be the set of all continuous real-valued functions of s on the interval $0 \leqslant s \leqslant 1$. Let E be the subset of X consisting of those functions x for which $x(0) = x(1)$. Then

$$E = \{x:x(0) = x(1)\}.$$

Example 2. Let X be the set of all real-valued functions x which are defined and have derivatives of all orders for all real values of s. Let E be the set of those x such that x and all its derivatives have the value 0 at $s = 0$. Then

$$E = \{x:x^{(n)}(0) = 0, \qquad n = 0, 1, 2, \cdots\}.$$

Subsets of a given set can be combined in two ways: by the formation of *unions* and *intersections.* If $E_1, \cdots, E_n$ are subsets of X, the union of $E_1, \cdots, E_n$ is defined as the set of all elements of X which belong to *at least* one of the sets $E_1, \cdots, E_n$. This union is denoted by

$$E_1 \cup E_2 \cup \cdots \cup E_n, \qquad \text{or} \qquad \bigcup_{i=1}^{n} E_i.$$

This definition and notation are extended to arbitrary collections of subsets of X as follows: Suppose $\mathscr{E}$ is a collection (family) of subsets of X. The set of all $x \in X$ such that $x \in E$ for at least one $E \in \mathscr{E}$ is denoted by

$$\bigcup_{E \in \mathscr{E}} E$$

and is called the union of the sets of $\mathscr{E}$. In case $\mathscr{E}$ is a countable set, with members $E_1, E_2, \cdots$, we write the union as

$$\bigcup_{n=1}^{\infty} E_n.$$

If the members of $\mathscr{E}$ are indexed in some manner, say E_α, where α ranges over a set A, we may denote the union by

$$\bigcup_{\alpha \in A} E_\alpha.$$

The *intersection* of a collection $\mathscr{E}$ of subsets E of X is defined as the set of all $x \in X$ such that $x \in E$ for *every* $E \in \mathscr{E}$. It is denoted by

$$\bigcap_{E \in \mathscr{E}} E$$

If the members of $\mathscr{E}$ are indexed, the intersection is represented symbolically in the appropriate manner corresponding to the usage for unions, but with

$$\bigcap \text{ in place of } \bigcup.$$

The intersection of the finite collection $E_1, \cdots, E_n$ is also written in the form

$$E_1 \cap E_2 \cap \cdots \cap E_n.$$

If $E \subset X$, the *complement* of E (relative to X) is defined as the set of all $x \in X$ such that $x \notin E$. We denote the complement of E by E', or, in some cases where the prime notation is inconvenient, by $C(E)$. Observe the following:

$$(E')' = E,$$

and if $E \subset F$, then $F' \subset E'$.

There is an important relation between complements, unions, and intersections; namely: If $\mathscr{E}$ is a collection of subsets of X, the complement of the union of all the members of $\mathscr{E}$ is the intersection of their complements. That is,

$$\left(\bigcup_{E \in \mathscr{E}} E\right)' = \bigcap_{E \in \mathscr{E}} E'.$$

For a finite collection this can be written

$$(E_1 \cup E_2 \cup \cdots \cup E_n)' = E_1' \cap E_2' \cap \cdots \cap E_n'.$$

If we replace E_i by E_i' throughout and then take the complement of both sides, we obtain the equivalent relation

$$(E_1 \cap E_2 \cap \cdots \cap E_n)' = E_1' \cup E_2' \cup \cdots \cup E_n'.$$

The general form of this is

$$\left(\bigcap_{E \in \mathscr{E}} E\right)' = \bigcup_{E \in \mathscr{E}} E'.$$

If E and F are subsets of X, the *difference* $E - F$ is defined as the set of all $x \in X$ which are in E but not in F. In other words,

$$E - F = E \cap F'.$$

Functions

In this book, unless we make special mention to the contrary, we always use the word "function" to mean a single-valued function. Since we shall have a great deal to do with functions whose arguments and values need not be real or complex numbers but may be elements of quite arbitrary sets, we put down here explicitly our general definition of a function.

Let X and Y be arbitrary nonempty sets. Suppose there is some rule whereby to each element $x \in X$ corresponds a uniquely determined element $y \in Y$. Consider the set consisting of all ordered pairs (x, y), where $x \in X$ and y is the corresponding element of Y. This set of pairs is called a *function.* The set X is called the *domain* of the function. The set of y's which occur as correspondents of the x's may or may not comprise all of Y. In any event, this set of y's is called the *range* of the function.

We see by this definition that a function with domain X and range contained in the set Y is a subset of the set of *all* ordered pairs (x, y) with $x \in X$ and $y \in Y$. This latter set of *all* such pairs is denoted by $X \times Y$ and called the *Cartesian product* (or often, just the *product*) of X and Y. Thus, a function with domain X and range in Y is a particular kind of subset of $X \times Y$. The distinguishing property of a function, as contrasted with other subsets of $X \times Y$, is that among the pairs (x, y) which form the distinct elements of the function, *each x in X occurs once and only once.* Any subset of $X \times Y$ with this property is a function with domain X and range in Y. Thus, let F be a subset of $X \times Y$. Then F is a function with domain X and range in Y if and only if the two following conditions are satisfied:

1. To each $x \in X$ corresponds some $y \in Y$ such that $(x, y) \in F$,
2. If (x_1, y_1) and (x_2, y_2) are in F and $y_1 \neq y_2$, then $x_1 \neq x_2$.

Sometimes we have occasion to consider nonempty sets X, Y and a function whose domain is a proper subset of X and whose range is contained in Y. In this case the function is a subset of $X \times Y$ which satisfies the second but not the first of the two foregoing conditions.

If F is a function and $(x, y) \in F$, we write $y = F(x)$; this is the usual functional notation. In general, we find it best to refer to the function by the single letter F, reserving the notation $F(x)$ for the element of the range of F which corresponds to the element x of the domain. Sometimes, however, it is convenient to do otherwise. Thus, in speaking of the exponential function, we may refer to the function e^x, or the function $y = e^x$, instead of using the lengthy phraseology "the function consisting of all pairs (x, y), where $y = e^x$ and x varies over all real numbers."

Inverse Functions

Suppose F is a function with domain D and range R, where $D \subset X$ and $R \subset Y$. Consider the Cartesian product $Y \times X$; note the reversal of order. Consider the subset of $Y \times X$ consisting of those elements (y, x) such that $(x, y) \in F$. It *may* be that this subset of $Y \times X$ is a function (with domain in Y and range in X). If so, we call it the function *inverse* to F, and we often denote it by F^{-1}. The domain of F^{-1} is then R, and its range is D.

Observe that the inverse of F is defined if and only if the correspondence between x and $F(x)$, as x varies over D, is a one-to-one correspondence between the elements of D and the elements of R. Another way of putting it is that F has an inverse if and only if $F(x_1) = F(x_2)$ implies that $x_1 = x_2$.

Also observe that, when F has an inverse, $y = F(x)$ is equivalent to $x = F^{-1}(y)$. Finally, if F has the inverse F^{-1}, then F^{-1} also has an inverse, namely F.

Usage Relating to Real Numbers

The least upper bound (if it exists) of a set S of real numbers is denoted by sup S. The greatest lower bound of S is denoted by inf S. The sup and inf notation is also used in other appropriate situations. For instance, if f is a real-valued function whose domain includes the set E, the least upper bound of the set of all $f(x)$ corresponding to x in E is denoted by

$$\sup_{x \in E} f(x).$$

We follow standard usage with respect to the symbols $+\infty$, $-\infty$ in relation to the real number system.

Inequalities

At a number of places in this book we use some of the standard inequalities concerning sums and integrals. We list the most commonly used ones here. The standard reference work on this subject is the book, *Inequalities*, by Hardy, Littlewood, and Pólya. In what follows we refer to this book as H, L, and P, and cite by number the section in which the stated inequality is discussed. In all inequalities the quantities involved

Functions

In this book, unless we make special mention to the contrary, we always use the word "function" to mean a single-valued function. Since we shall have a great deal to do with functions whose arguments and values need not be real or complex numbers but may be elements of quite arbitrary sets, we put down here explicitly our general definition of a function.

Let X and Y be arbitrary nonempty sets. Suppose there is some rule whereby to each element $x \in X$ corresponds a uniquely determined element $y \in Y$. Consider the set consisting of all ordered pairs (x, y), where $x \in X$ and y is the corresponding element of Y. This set of pairs is called a *function.* The set X is called the *domain* of the function. The set of y's which occur as correspondents of the x's may or may not comprise all of Y. In any event, this set of y's is called the *range* of the function.

We see by this definition that a function with domain X and range contained in the set Y is a subset of the set of *all* ordered pairs (x, y) with $x \in X$ and $y \in Y$. This latter set of *all* such pairs is denoted by $X \times Y$ and called the *Cartesian product* (or often, just the *product*) of X and Y. Thus, a function with domain X and range in Y is a particular kind of subset of $X \times Y$. The distinguishing property of a function, as contrasted with other subsets of $X \times Y$, is that among the pairs (x, y) which form the distinct elements of the function, *each x in X occurs once and only once.* Any subset of $X \times Y$ with this property is a function with domain X and range in Y. Thus, let F be a subset of $X \times Y$. Then F is a function with domain X and range in Y if and only if the two following conditions are satisfied:

1. To each $x \in X$ corresponds some $y \in Y$ such that $(x, y) \in F$,
2. If (x_1, y_1) and (x_2, y_2) are in F and $y_1 \neq y_2$, then $x_1 \neq x_2$.

Sometimes we have occasion to consider nonempty sets X, Y and a function whose domain is a proper subset of X and whose range is contained in Y. In this case the function is a subset of $X \times Y$ which satisfies the second but not the first of the two foregoing conditions.

If F is a function and $(x, y) \in F$, we write $y = F(x)$; this is the usual functional notation. In general, we find it best to refer to the function by the single letter F, reserving the notation $F(x)$ for the element of the range of F which corresponds to the element x of the domain. Sometimes, however, it is convenient to do otherwise. Thus, in speaking of the exponential function, we may refer to the function e^x, or the function $y = e^x$, instead of using the lengthy phraseology "the function consisting of all pairs (x, y), where $y = e^x$ and x varies over all real numbers."

Inverse Functions

Suppose F is a function with domain D and range R, where $D \subset X$ and $R \subset Y$. Consider the Cartesian product $Y \times X$; note the reversal of order. Consider the subset of $Y \times X$ consisting of those elements (y, x) such that $(x, y) \in F$. It *may* be that this subset of $Y \times X$ is a function (with domain in Y and range in X). If so, we call it the function *inverse* to F, and we often denote it by F^{-1}. The domain of F^{-1} is then R, and its range is D.

Observe that the inverse of F is defined if and only if the correspondence between x and $F(x)$, as x varies over D, is a one-to-one correspondence between the elements of D and the elements of R. Another way of putting it is that F has an inverse if and only if $F(x_1) = F(x_2)$ implies that $x_1 = x_2$.

Also observe that, when F has an inverse, $y = F(x)$ is equivalent to $x = F^{-1}(y)$. Finally, if F has the inverse F^{-1}, then F^{-1} also has an inverse, namely F.

Usage Relating to Real Numbers

The least upper bound (if it exists) of a set S of real numbers is denoted by sup S. The greatest lower bound of S is denoted by inf S. The sup and inf notation is also used in other appropriate situations. For instance, if f is a real-valued function whose domain includes the set E, the least upper bound of the set of all $f(x)$ corresponding to x in E is denoted by

$$\sup_{x \in E} f(x).$$

We follow standard usage with respect to the symbols $+\infty$, $-\infty$ in relation to the real number system.

Inequalities

At a number of places in this book we use some of the standard inequalities concerning sums and integrals. We list the most commonly used ones here. The standard reference work on this subject is the book, *Inequalities*, by Hardy, Littlewood, and Pólya. In what follows we refer to this book as H, L, and P, and cite by number the section in which the stated inequality is discussed. In all inequalities the quantities involved may be either real or complex. Sums are either all from 1 to n or from 1 to ∞, and in the latter case certain evident assumptions and implications of convergence are involved. For simplicity the inequalities for integrals are written for the case in which the functions are defined on a finite or infinite interval of the real axis. The inequalities are valid with more general interpretations of the set over which integration is extended.

Hölder's inequality for sums (H, L, and P, § 2.8): If $1 < p < \infty$ and $p' = \dfrac{p}{p-1}$, then

$$\sum |a_i b_i| \leqslant \left(\sum |a_i|^p\right)^{1/p} \left(\sum |b_i|^{p'}\right)^{1/p'}.$$

The special case when $p = p' = 2$ is called Cauchy's inequality (H, L, and P, § 2.4).

Minkowski's inequality for sums (H, L, and P, § 2.11): If $1 \leqslant p < \infty$, then

$$\left(\sum |a_i + b_i|^p\right)^{1/p} \leqslant \left(\sum |a_i|^p\right)^{1/p} + \left(\sum |b_i|^p\right)^{1/p}.$$

Jensen's inequality (H, L, and P, § 2.10): If $0 < p < q$, then

$$\left(\sum |a_i|^q\right)^{1/q} \leqslant \left(\sum |a_i|^p\right)^{1/p}.$$

Hölder's inequality for integrals (H, L, and P, § 6.9): If $1 < p < \infty$ and $p' = \dfrac{p}{p-1}$, then

$$\int |f(x)\, g(x)|\, dx \leqslant \left(\int |f(x)|^p\, dx\right)^{1/p} \left(\int |g(x)|^{p'}\, dx\right)^{1/p'}.$$

The special case when $p = p' = 2$ is called the Schwarz inequality (H, L, and P, § 6.5).

Minkowski's inequality for integrals (H, L, and P, § 6.13): If $1 \leqslant p < \infty$, then

$$\left(\int |f(x) + g(x)|^p\, dx\right)^{1/p} \leqslant \left(\int |f(x)|^p\, dx\right)^{1/p} + \left(\int |g(x)|^p\, dx\right)^{1/p}.$$

The Kronecker Delta

The symbol δ_{ij} denotes the number 1 if $i = j$ and the number 0 if $i \neq j$. Here i and j are positive integers.

1

THE ABSTRACT APPROACH TO LINEAR PROBLEMS

1.0 Introductory Statement

The modern treatment of many topics in pure and applied mathematics is characterized by the effort which is made to strip away nonessential details and to show clearly the fundamental assumptions and the structure of the reasoning. This effort often leads to a certain degree of abstraction, the concrete nature of the originally contemplated problem being temporarily put aside, and the aspects of the problem which are of greatest significance being cast into axiomatic form. It is found that in this way there is a considerable gain in transparency and that diverse problems exhibit common characteristics which enable them all to be at least partially solved by the methods of a single general theory.

In this chapter we consider the algebraic aspects of such an abstract approach to linear problems. In essence, all linear problems are viewed in some measure as analogous to the linear problems exhibited in elementary algebra by the theory of systems of linear equations. The linear problems of analysis usually require topological as well as algebraic considerations. However, in this chapter, we exclude all concern with topology; the topological aspects of the abstract approach to linear problems will be taken up in later chapters.

The profoundest results of the chapter are the extension theorems in § 1.71 (Theorems 1.71–A and 1.71–D). They depend on Zorn's lemma. Theorem 1.71–A is needed to demonstrate the existence of certain projections (see Theorem 4.8–A). Theorem 1.71–D is the fundamental tool in proofs concerning the extension of continuous linear functionals and, in particular, in the proof of the Hahn-Banach theorem (see Theorems 3.7–B, 3.8–D, and 4.3–A).

Chapter 1 culminates in § 1.91 with two groups of theorems: one group (the first three theorems) concerning the range of a linear operator and the null manifold of the transpose of the operator and the other group (the last four theorems) concerning the null manifold of the operator and the range of its transpose. These theorems furnish information on existence and uniqueness theorems in the case of certain kinds of linear problems. For the finite-dimensional case these theorems include the standard results concerning algebraic systems of linear systems. In the infinite-dimensional case the results are not as useful as results which can be obtained with the aid of metric or topological tools. Nevertheless, the material of § 1.91 points the way to more incisive results, some of which are given in § 4.6.

1.1 Abstract Linear Spaces

We have as yet made no formal definition of what is meant by the adjective *linear* in the phrase "linear problems." We can cite various particular kinds of linear problems: the problems of homogeneous and inhomogeneous systems of linear equations in n "unknowns" in elementary algebra; the problems of the theory of linear ordinary differential equations (existence theorems, particular and general solutions, problems of finding solutions satisfying given conditions at one or two end points); boundary or initial-value problems in the theory of linear partial differential equations; problems in the theory of linear integral equations; linear "transform" problems, e.g., problems related to Fourier and Laplace transforms. This is by no means an exhaustive list of the types of mathematical situation in which linear problems arise.

At the bottom of every linear problem is a mathematical structure called a *linear space*. We shall therefore begin with an axiomatic treatment of abstract linear spaces.

A collection of elements, together with a certain structure of relations between elements or of rules of manipulation and combination, the whole supporting a mathematical development, is often called a *space*. This terminology derives from the model of geometry, in which the elements are points. We are now going to define what is meant, abstractly, by a *linear space.*

Definition. Let X be a set of elements, hereafter sometimes called *points*, and denoted by small italic letters: $x, y, \cdots$. We assume that each pair of elements x, y can be combined by a process called addition to yield another element z denoted by $z = x + y$. We also assume that each real number α and each element x can be combined by a process called multiplication to yield another element y denoted by $y = \alpha x$. The set X with these two processes is called a *linear space* if the following axioms are satisfied:

1. $x + y = y + x$.
2. $x + (y + z) = (x + y) + z$.
3. There is in X a unique element, denoted by 0 and called the zero element, such that $x + 0 = x$ for each x.
4. To each x in X corresponds a unique element, denoted by $-x$, such that $x + (-x) = 0$.
5. $\alpha(x + y) = \alpha x + \alpha y$.
6. $(\alpha + \beta)x = \alpha x + \beta x$.
7. $\alpha(\beta x) = (\alpha\beta)x$.
8. $1 \cdot x = x$.
9. $0 \cdot x = 0$.

Anyone who is familiar with the algebra of vectors in ordinary three-dimensional Euclidean space will see at once that the set of all such vectors forms a linear space. An abstract linear space embodies so many of the features of ordinary vector algebra that the word *vector* has been taken over into a more general context. A linear space is often called a *vector space*, and the elements of the space are called *vectors.*

In the foregoing list of axioms it was assumed that the multiplication operation was performed with *real* numbers α, β. To emphasize this, if necessary, we call the space a *real* linear space, or a *real* vector space. An alternative notion of a linear space is obtained if it is assumed that any *complex* number α and any element x can be multiplied, yielding another element αx. The axioms are the same as before. The space is then called a *complex* linear space.

The notion of a vector space is defined even more generally in abstract algebra, by allowing the multipliers $\alpha, \beta, \cdots$ to be elements of an arbitrary commutative field. In this book, however, we confine ourselves to the two fields of real and complex numbers, respectively. The elements of the field are called *scalars*, to contrast with the *vector* elements of the linear space.

It is easy to see that $-1 \cdot x = -x$ and that $\alpha \cdot 0 = 0$. We write $x - y$ for convenience in place of $x + (-y)$. The following "cancellation" rules are also easily deduced from the axioms:

(1.1–1) $x + y = x + z$ implies $y = z$;

(1.1–2) $\alpha x = \alpha y$ and $\alpha \neq 0$ imply $x = y$;

(1.1–3) $\alpha x = \beta x$ and $x \neq 0$ imply $\alpha = \beta$.

With respect to addition, a linear space X is an Abelian (commutative) *group*, in the technical algebraic sense.

Definition. A nonempty subset M of a linear space X is called a *linear manifold* in X if $x + y$ is in M whenever x and y are both in M and if also αx is in M whenever x is in M and α is any scalar.

In this definition and generally throughout the book, statements made about linear spaces, without qualification as to whether the space is real or complex, will be intended to apply equally to real spaces and complex spaces.

It will be seen at once that, if M is a linear manifold in X, it may be regarded as a linear space by itself. For, if x is in M, then $-1 \cdot x = -x$ is also in M, and $x - x = 0$ is also in M. The nine axioms for a linear space are now found to be satisfied in M. Another term for a linear manifold in X is *subspace* of X. A subspace of X is called *proper* if it is not all of X.

The set consisting of 0 alone is a subspace. We denote it by (0).

Suppose S is any nonempty subset of X. Consider the set M of all finite linear combinations of elements of S, i.e., elements of the form $\alpha_1 x_1 + \cdots + \alpha_n x_n$, where n is any positive integer (not fixed), $x_1, \cdots, x_n$ are any elements of S, and $\alpha_1, \cdots, \alpha_n$ are any scalars. This set M is a linear manifold. It is called the linear manifold generated, or determined, by S. Sometimes we speak of M as the linear manifold *spanned* by S. It is easy to verify the truth of the following statements: (1) M consists of those vectors which belong to every linear manifold which contains S; i.e., M is the intersection of all such manifolds. (2) M is the smallest linear manifold which contains S; i.e., if N is a linear manifold which contains S, then M is contained in N.

One of the most important concepts in a vector space is that of linear dependence.

Definition. A finite set of vectors $x_1, \cdots, x_n$ in the space X is *linearly dependent* if there exist scalars $\alpha_1, \cdots, \alpha_n$, not all zero, such that $\alpha_1 x_1 + \cdots + \alpha_n x_n = 0$. If the finite set $x_1, \cdots, x_n$ is not linearly dependent, it is called *linearly independent.* In that case, a relation $\alpha_1 x_1 + \cdots + \alpha_n x_n = 0$ implies that $\alpha_1 = \cdots = \alpha_n = 0$. An infinite set S of vectors is called linearly independent if every finite subset of S is linearly independent; otherwise S is called linearly dependent.

Observe that, if a set of vectors contains a linearly dependent subset, the whole set is linearly dependent. Also note that a linearly independent set cannot contain the vector 0.

We note the following simple but important theorem, of which use will be made in later arguments:

Theorem 1.1–A. *Suppose $x_1, \cdots, x_n$ is a set of vectors with $x_1 \neq 0$. The set is linearly dependent if and only if some one of the vectors $x_2, \cdots, x_n$, say x_k, is in the linear manifold generated by $x_1, \cdots, x_{k-1}$.*

PROOF. Suppose the set is linearly dependent. There is a smallest integer k, with $2 \leqslant k \leqslant n$, such that the set $x_1, \cdots, x_k$ is linearly dependent. This dependence is expressed by an equation $\alpha_1 x_1 + \cdots + \alpha_k x_k = 0$, with not all the α's equal to zero. Necessarily, then, $\alpha_k \neq 0$, for otherwise $x_1, \cdots, x_{k-1}$ would form a linearly dependent set. Consequently $x_k = \beta_1 x_1 + \cdots + \beta_{k-1} x_{k-1}$, where $\beta_i = -\alpha_i/\alpha_k$. This shows that x_k is in the manifold spanned by $x_1, \cdots, x_{k-1}$. On the other hand, if we assume that some x_k is in the linear manifold spanned by $x_1, \cdots, x_{k-1}$, then an equation of the form $x_k = \beta_1 x_1 + \cdots + \beta_{k-1} x_{k-1}$ shows that the set $x_1, \cdots, x_k$ is linearly dependent, whence the same is true of the set $x_1, \cdots, x_n$.

It is convenient to say that x is a *linear combination* of $x_1, \cdots, x_n$ if it is in the linear manifold spanned by these vectors.

Using the notion of linear dependence, we can define the concept of a finite-dimensional vector space.

Definition. Let X be a vector space. Suppose there is some positive integer n such that X contains a set of n vectors which are linearly independent, while every set of $n + 1$ vectors in X is linearly dependent. Then X is called *finite dimensional*, and n is called the dimension of X. A vector space with just one element (which must then be the zero element) is also called finite dimensional, of dimension zero. If X is not finite dimensional, it is called infinite dimensional.

As we shall see later, the spaces of greatest interest in analysis are infinite dimensional. Nevertheless, it will often be of use to consider finite-dimensional spaces. Such spaces are, moreover, the source of much of our intuitive perception about what to expect in dealing with linear spaces generally.

Definition. A finite set S in a space X is called a *basis* of X if S is linearly independent and if the linear manifold generated by S is all of X.

If $x_1, \cdots, x_n$ is a basis of X, the definition means that every x in X can be expressed in the form $x = \xi_1 x_1 + \cdots + \xi_n x_n$. Since the basis is a linearly independent set, the coefficients $\xi_1, \cdots, \xi_n$ are uniquely determined by x; that is, x cannot be expressed as a *different* linear combination of the basis elements.

It is readily seen that, if X is n-dimensional, where $n \geqslant 1$, then X has a basis consisting of n elements. For, X certainly contains vectors $x_1, \cdots, x_n$ which form a linearly independent set. Now, for any x, the set of $n + 1$ vectors $x_1, x_2, \cdots, x_n, x$ must be linearly dependent, by the definition of the dimensionality of X. Hence it is clear, by Theorem 1.1–A, that x is in the linear manifold spanned by $x_1, \cdots, x_n$. This shows that $x_1, \cdots, x_n$ form a basis of X.

Next we wish to show that if X has a basis of n elements, then X is n-dimensional. First we prove a lemma.

Lemma 1.1–B. *If the finite set $x_1, \cdots, x_n$ generates X, and if $y_1, \cdots, y_m$ are elements of X forming a linearly independent set, then $m \leqslant n$.*

PROOF. If S is any linearly dependent finite ordered set of vectors $u_1, \cdots, u_p$ with $u_1 \neq 0$, let S' denote the ordered set which remains after the deletion of the first u_i which is a linear combination of its predecessors. Also, for any y, let yS denote the ordered set $(y, u_1, \cdots, u_p)$. Now define S_1 to be the set $(y_m, x_1, \cdots, x_n)$, $S_2 = y_{m-1}S_1'$, $S_3 = y_{m-2}S_2'$, and so on. We make several observations: (1) S_1 spans X, is a linearly dependent set, and $y_m \neq 0$. Hence we can form S_1' (see Theorem 1.1–A). (2) S_1' and hence also S_2, spans X. We can continue in this way, constructing new sets S and S' as long as the y's last. Since the set of y's is linearly independent, the discarded element at each step must be an x. Since we can form $S_1', \cdots, S_m'$, it follows that we discard m x's, and hence $n \geqslant m$, as asserted in the lemma.

Theorem 1.1–C. *If the linear space X has a basis of n elements, X is n-dimensional, and conversely.*

PROOF. The proof of the converse has already been given (just before Lemma 1.1–B). If X has a basis of n elements, any linearly independent set in X has at most n elements, by Lemma 1.1–B. On the other hand, the basis is a set of n linearly independent elements. Hence X is n-dimensional, by definition.

The following theorem will be useful later on:

Theorem 1.1–D. *Let X be an n-dimensional vector space, and let the* *set $y_1, \cdots, y_m$ be linearly independent, with $m < n$. Then there exists a basis of X composed of $y_1, \cdots, y_m$ and $n - m$ other vectors.*

PROOF. Let $x_1, \cdots, x_n$ be a basis of X. Let S_1 be the ordered set $(y_1, \cdots, y_m, x_1, \cdots, x_n)$, and let $S_2 = S_1'$, $S_3 = S_2'$, and so on, where S' is related to S in the manner explained in the proof of Lemma 1.1–B. Observe that S_1 is certainly linearly dependent and that the deleted vector at each step is one of the x's (by Theorem 1.1–A), so that each S_k includes $y_1, \cdots, y_m$. Since S_1 spans X, so does S_1', and so on. We can form S_k' if and only if S_k is linearly dependent. Ultimately we reach a stage where S_k is linearly independent and spans X. It is then a basis of X and includes $y_1, \cdots, y_m$. Since the basis must have n elements, there are $n - m$ vectors in addition to $y_1, \cdots, y_m$.

It is natural to expect that, if X is a linear space of dimension n, every subspace of X has dimension not exceeding n. This is indeed the case.

Theorem 1.1–E. *Suppose the vector space X is of dimension n, and let M be a proper subspace of X. Then M is of some finite dimension m, where $m < n$.*

PROOF. If $M = (0)$, it is of dimension 0, by definition. Since X must contain nonzero elements, $n > 0$, so the assertion of the theorem is true in this case. We now assume that M contains nonzero elements. If x_0 is a nonzero element of M, it forms a linearly independent set in M. On the other hand, a linearly independent set in M cannot contain as many as n elements. For, a linearly independent set of n elements is a basis of X (as we saw following the definition of a basis), so that, in such a case X would be generated by M; this is impossible, since $M \neq X$, and the linear manifold generated by the manifold M is M itself. We now consider the nonempty class of all linearly independent sets in M. Each such set has a finite number of elements p, where $1 \leqslant p < n$. There is therefore a set for which p is largest, say $p = m$. It is immediate that $m < n$ and that M is of dimension m.

Before going on to consider the general form of linear problems in linear spaces, we illustrate to some extent the great variety of possible examples of linear spaces.

1.2 Examples of Linear Spaces

Example 1. The simplest important example of a real linear space is the set of all n-tuples of real numbers, $x = (\xi_1, \cdots, \xi_n)$. The definitions of addition and multiplication by scalars in this space are as follows:

If $x = (\xi_1, \cdots, \xi_n)$ and $y = (\eta_1, \cdots, \eta_n)$, then $z = x + y$, where

$z = (\zeta_1, \cdots, \zeta_n)$, $\zeta_k = \xi_k + \eta_k$, $k = 1, \cdots, n$. The vector αx is $(\alpha\xi_1, \cdots, \alpha\xi_n)$. Here the ξ's, η's, and α are arbitrary real numbers, and n is an arbitrary, fixed positive integer. We define $0 = (0, \cdots, 0)$ and $-x = (-\xi_1, \cdots, -\xi_n)$. It is an easy matter to verify that the nine axioms for a linear space are satisfied.

The dimension of this space is n; we prove this by exhibiting a basis consisting of n elements. Let $e_1 = (1, 0, \cdots, 0)$, $e_2 = (0, 1, 0, \cdots, 0)$, $\cdots$, $e_n = (0, 0, \cdots, 0, 1)$. If $x = (\xi_1, \cdots, \xi_n)$, observe that $x = \xi_1 e_1 + \cdots + \xi_n e_n$. Thus, the set $e_1, \cdots, e_n$ generates the whole space. Moreover, the set is linearly independent, for $\alpha_1 e_1 + \cdots + \alpha_n e_n = (\alpha_1, \cdots, \alpha_n) = 0$ if and only if all the α's are zero. Thus $e_1, \cdots, e_n$ constitute a basis.

We call this space *n-dimensional real arithmetic space*, and denote it by R_n.

The space R_n has a familiar geometrical interpretation, $\xi_1, \cdots, \xi_n$ being the coordinates of the point x in a system of Cartesian coordinates. Thus, R_1 is interpreted as a line, R_2 as a plane, and so on. In the geometrical interpretation of R_n we may regard the element x either as a point or as the vector from 0 (the origin in R_n) to that point. For geometrical interpretations this latter point of view is in many ways the most fruitful.

Example 2. The set of all n-tuples $x = (\xi_1, \cdots, \xi_n)$ of *complex* numbers forms a complex linear space, which we call *n-dimensional complex arithmetic space*, and denote by C_n.

The n vectors $e_1 = (1, 0, \cdots, 0), \cdots, e_n = (0, \cdots, 0, 1)$ form a basis of C_n. It will be observed that the elements of R_n belong to C_n; however, R_n is not a subspace of C_n, for, if α is complex and x is in R_n, then αx is in C_n but not always in R_n (e.g., $i \cdot e_1 = (i, 0, \cdots, 0)$ is not in R_n).

The set C_n can also be regarded as a *real* linear space, by using the real field for scalar multipliers. But then the space is not of dimension n, but of dimension $2n$, one possible basis consisting of the vectors $e_1, \cdots, e_n$ and $ie_1, \cdots, ie_n$. We shall hereafter always regard C_n as a *complex* space.

Example 3. Let $[a, b]$ (with $a < b$) be a finite closed interval of the real axis, and let x denote a continuous real-valued function whose value at the point s of $[a, b]$ is $x(s)$. Let $C[a, b]$ denote the set of all such functions, and define $x_1 + x_2$, αx in the natural manner, i.e., $z = x_1 + x_2$, where the value of z at s is $z(s) = x_1(s) + x_2(s)$; $y = \alpha x$, where the value of y at s is $y(s) = \alpha x(s)$. It is clear that $C[a, b]$ is a real vector space.

We might equally well have considered complex-valued continuous functions of s; in that case we would have obtained a complex vector space. We shall denote this space by $C[a, b]$ also; thus, in speaking of the space $C[a, b]$, we shall have to make clear by an explicit statement whether we are talking about real-valued or complex-valued functions. In either case, $C[a, b]$ is infinite dimensional. For, let $x_0(s) = 1$, $x_n(s) = s^n$, $n = 1, 2, \cdots$. Evidently, $x_0, x_1, \cdots, x_n$ all belong to $C[a, b]$. This set of elements is linearly independent, no matter how large n is; for, by well known properties of polynomials, if $\alpha_0 + \alpha_1 s + \cdots + \alpha_n s^n = 0$ for every s such that $a \leqslant s \leqslant b$, then $\alpha_0 = \alpha_1 = \cdots = \alpha_n = 0$. Therefore $C[a, b]$ cannot be finite dimensional.

The interval $[a, b]$ does not play a very important role in the demonstration that $C[a, b]$ is a vector space and of infinite dimension. For example, we could equally well have considered continuous functions of a complex variable which ranges over some fixed infinite point set in the complex plane. An infinite set is essential to make the space infinite dimensional.

A great many of the linear spaces which are of interest in analysis are spaces whose elements are functions. We shall mention a few further examples.

Example 4. Let f be a function of the complex variable z which is defined (single valued), analytic, and bounded in the open unit circle $|z| < 1$. The class of all such functions becomes a complex vector space when $f + g$ and αf are defined in the natural way. This space is infinite dimensional. As a subspace we mention the class of all those functions f in the space for which $f(0) = 0$. Another subspace is the class of those f's whose definitions can be extended to the boundary $|z| = 1$ in such a way that each f is continuous in the closed circle $|z| \leqslant 1$. Both of these subspaces are infinite dimensional.

Example 5. Consider the complex-valued functions x of the real variable s which are such that $x(s)$, $x'(s)$ and $x''(s)$ are defined and continuous on the closed interval $[0, \pi]$. The set of all such functions is a linear space. It is of interest in considering ordinary second-order linear differential equations with coefficients continuous on $[0, \pi]$. The set of all elements of the space which satisfy the equation $x''(s) + x(s) = 0$ is a subspace of dimension 2. A basis of the subspace is furnished by the functions e^{is}, e^{-is}. Another basis is furnished by $\sin s$ and $\cos s$. Another subspace of interest is that consisting of all elements of the space such that $x(0) = x(\pi) = 0$. This subspace is infinite dimensional, for it contains the infinite linearly independent set consisting of $\sin ns$, $n = 1, 2, \cdots$.

Example 6. As elements of a space consider infinite sequences $x = \{\xi_n\}$ $(n = 1, 2, \cdots)$ such that $\sum_{n=1}^{\infty} |\xi_n|^2 < \infty$, the ξ's being complex numbers. Define $\alpha x = \{\alpha\xi_n\}$ and $\{\xi_n\} + \{\eta_n\} = \{\xi_n + \eta_n\}$. It is readily seen (by Minkowski's inequality with $p = 2$) that this is an infinite-dimensional complex linear space. We denote it by l^2. It is the space which was first extensively studied by D. Hilbert in his work on quadratic forms in infinitely many variables, with applications to the theory of integral equations. On this account l^2 is the classical prototype of the variety of linear space known today under the name *Hilbert* space. Hilbert spaces are linear spaces with a certain special kind of metrical structure. We discuss them extensively elsewhere in this book.

Example 7. Suppose $p \geqslant 1$ (p not necessarily an integer). Let $\mathscr{L}^p$ denote the class of all functions x of the real variable s such that $x(s)$ is defined for all s, with the possible exception of a set of measure zero, and is measurable and $|x(s)|^p$ is integrable (in the Lebesgue sense) over the range $(-\infty, \infty)$. Instead of $(-\infty, \infty)$ we could equally well consider $(0, \infty)$ or any finite interval (a, b). We write $\mathscr{L}^p(-\infty, \infty)$, $\mathscr{L}^p(0, \infty)$, $\mathscr{L}^p(a, b)$ to distinguish these various situations. Also, we usually omit the index p when $p = 1$, writing $\mathscr{L}$ for $\mathscr{L}^1$.

Let $\mathscr{D}_x$ denote the set on which x is defined. We define αx as the function $(\alpha x)(s) = \alpha x(s)$, with $\mathscr{D}_{\alpha x} = \mathscr{D}_x$, and $x + y$ as the function $(x + y)(s) = x(s) + y(s)$, with $\mathscr{D}_{x+y} = \mathscr{D}_x \cap \mathscr{D}_y$. Clearly $\alpha x \in \mathscr{L}^p$ if $x \in \mathscr{L}^p$; it is also true that $x + y \in \mathscr{L}^p$ if $x, y \in \mathscr{L}^p$. This latter fact follows from the inequality

$$|A + B|^p \leqslant 2^p[|A|^p + |B|^p], \tag{1.2–1}$$

where A and B are any real or complex numbers. For, if w is measurable, if z is integrable, and if $|w(s)| \leqslant |z(x)|$, then w is integrable. To see the truth of (1.2–1), observe that

$$\max\{|A|, |B|\} \leqslant |A| + |B| \leqslant 2\max\{|A|, |B|\},$$

$$|A + B|^p \leqslant (|A| + |B|)^p \leqslant (2\max\{|A|, |B|\})^p$$
$$= \max\{2^p|A|^p, 2^p|B|^p\} \leqslant 2^p|A|^p + 2^p|B|^p.$$

If we define $x = y$ to mean that $\mathscr{D}_x = \mathscr{D}_y$ and $x(s) = y(s)$ for every $s \in \mathscr{D}_x$, we may at first think that $\mathscr{L}^p$ is a linear space. It is not, however. For, if $\mathscr{L}^p$ were a linear space, the zero element z would necessarily be the function defined for *all* s, with $z(s) = 0$ for every s. But then $x + (-x) = z$ would not be true for all x, as we see by choosing an x for which $\mathscr{D}_x$ does not include every s. To get around this difficulty we proceed as follows: Define an equivalence relation $=^0$ in $\mathscr{L}^p$ by saying that $x =^0 y$ if $x(s) = y(s)$ a.e. (i.e., almost everywhere, which means except on a set of measure zero). The set of equivalence classes into which $\mathscr{L}^p$ is thus divided is denoted by L^p; here also, we write L for L^1. For the time being we shall denote an element of L^p which contains x by $[x]$. We define $[x] + [y] = [x + y]$. This definition of $[x] + [y]$ is unambiguous, for, if $x =^0 x_1$ and $y =^0 y_1$, it follows that $x_1 + y_1 =^0 x + y$. Likewise, we define $\alpha[x] = [\alpha x]$, noting that $\alpha x =^0 \alpha x_1$ if $x =^0 x_1$. With these definitions the class L^p becomes a linear space; the zero element of L^p is the equivalence class consisting of all $x \in \mathscr{L}^p$ such that $x(s) = 0$ a.e.

In practice we usually ignore the notational distinction between $\mathscr{L}^p$ and L^p and write x instead of $[x]$. When we do this it must be remembered that x really denotes, not a single function, but a class of equivalent functions.

The space L^p is of interest in connection with Fourier transforms and various kinds of integral equation problems. The case $p = 2$ is especially important.

As an example of a subclass of $\mathscr{L}^p$ that is useful in certain applications to first-order ordinary differential equation problems, we mention the following: the class of all x for which $x(s)$ is defined everywhere and is absolutely continuous on every finite interval and both $x(s)$ and $x'(s)$ define elements of $\mathscr{L}^p$.

Example 8. Let $[a, b]$ be a finite closed interval of the real axis. Let $BV[a, b]$ denote the class of real-valued functions of s which are defined and of bounded variation on $[a, b]$. This class is a linear space.

Example 9. Let x be a complex-valued function of the real variable s which is defined and has derivatives of all orders for every value of s. The class of all such functions x is evidently a linear space.

It would be easy to give many more examples of linear spaces whose elements are certain kinds of functions. We shall see later on in this book that the study of linear spaces leads us to introduce still other linear spaces composed of functions defined on the original linear spaces.

1.3 Linear Operators

A linear operator is a certain kind of function whose domain is a linear space and whose range is contained in another linear space (possibly the same as the first one). For the terminology concerning functions we refer the reader to the Introduction. If A is a linear operator, it is customary to omit parentheses and write Ax instead of $A(x)$ whenever it seems convenient.

Definition. Let X and Y be linear spaces (both real or both complex). Let A be a function with domain X and range contained in Y. Then A is called a linear operator, or more explicitly a linear operator on X into Y, if the following two conditions are satisfied:

$$A(x_1 + x_2) = Ax_1 + Ax_2, \tag{1.3–1}$$

$$A(\alpha x) = \alpha Ax. \tag{1.3–2}$$

Here α is an arbitrary scalar, and x_1, x_2, x are arbitrary vectors from X.

It follows immediately by induction from (1.3–1) and (1.3–2) that

$$A(\alpha_1 x_1 + \cdots + \alpha_n x_n) = \alpha_1 Ax_1 + \cdots + \alpha_n Ax_n, \tag{1.3–3}$$

for arbitrary n. Also, by taking $\alpha = 0$ in (1.3–2), we see that

$$A(0) = 0. \tag{1.3–4}$$

Sometimes we consider a linear operator A whose domain is a proper subset of a given linear space X. The domain is itself a linear space (by definition), and so it is a subspace of X. Our standard notation for the domain of A is $\mathscr{D}(A)$. We denote the range of A by $\mathscr{R}(A)$.

Theorem 1.3–A. *Let A be a linear operator with domain $\mathscr{D}(A) \subset X$ and range $\mathscr{R}(A) \subset Y$, where X and Y are linear spaces. Then $\mathscr{R}(A)$ is a linear manifold in Y.*

PROOF. Suppose $y_1, y_2 \in \mathscr{R}(A)$, and let α be a scalar. We have to prove that $y_1 + y_2$ and αy_1 are in $\mathscr{R}(A)$. Now $y_1 \in \mathscr{R}(A)$ means that there is an $x_1 \in \mathscr{D}(A)$ such that $Ax_1 = y_1$. Likewise $Ax_2 = y_2$ for some $x_2 \in \mathscr{D}(A)$. Then, since A is linear, $A(x_1 + x_2) = Ax_1 + Ax_2 = y_1 + y_2$, and $A(\alpha x_1) = \alpha Ax_1 = \alpha y_1$. The desired conclusions are now evident.

If a linear operator A has an inverse, we denote it by A^{-1}. (The reader should consult the Introduction for the definition of the inverse of a function.) The statement "A^{-1} exists" means the same as "A has an inverse."

Theorem 1.3–B. *The inverse of a linear operator A exists if and only if $Ax = 0$ implies $x = 0$. When A^{-1} exists it is a linear operator.*

PROOF. We know by the definition of an inverse function that A^{-1} exists if and only if $Ax_1 = Ax_2$ implies $x_1 = x_2$. Suppose $Ax = 0$ implies $x = 0$, and let $Ax_1 = Ax_2$. Then, by the linearity of A, $A(x_1 - x_2) = 0$, whence $x_1 - x_2 = 0$, or $x_1 = x_2$. Thus A^{-1} exists. Now suppose, vice versa, that A^{-1} exists, and let x be a vector for which $Ax = 0$. Then $A(x) = A(0)$ [see (1.3–4)], and so, by the condition for the existence of A^{-1}, $x = 0$. This finishes the proof of the first statement in the theorem. The proof that A^{-1} is linear, when it exists, is left to the reader.

The linear problems of algebra and analysis are concerned with linear operators on various linear spaces. We mention two kinds of problems in very general terms: *existence* problems and *uniqueness* problems. Suppose A is a given linear operator, with domain a given space X and range in a given space Y. Then we can ask: "For which elements $y \in Y$ does there exist in X an element x such that $Ax = y$?" This is the same as asking: "Which elements of Y belong to $\mathscr{R}(A)$?" Existence problems are of this kind.

Example 1. Let $Y = C[0, 1]$ (real-valued functions). Let X be the subclass of Y consisting of those elements x which have first and second derivatives continuous on $[0, 1]$ and which are, moreover, such that $x(0) = x'(0) = 0$. Let p and q be members of Y, and define A on X by $Ax = y$ where

$$y(s) = x''(s) + p(s)x'(s) + q(s)x(s).$$

Then A is linear on X into Y. Is $\mathscr{R}(A)$ all of Y? This is the question as to whether the differential equation

$$x'' + px' + qx = y$$

has a solution such that $x \in X$, for each choice of $y \in Y$. Note that $x \in X$ implies that $x(0) = x'(0) = 0$. The "initial conditions" have been *incorporated into the definition of the domain of* A. There is a standard existence theorem in the theory of differential equations which does in fact assure us that $\mathscr{R}(A) = Y$. It also assures us that for a given y there is *only* one x in X such that $Ax = y$. Therefore A^{-1} exists.

Example 2. Let X be the space $BV[0, 1]$ defined in Example 8, § 1.2. Let Y be the space of all bounded sequences $y = \{\eta_i\}$ $(i = 1, 2, \cdots)$ where the definitions of the algebraic processes in Y are made as for the space l^2 of Example 6, § 1.2. Define a linear operator A on X into Y by $Ax = y$, where

$$\eta_k = \int_0^1 s^k \, dx(s).$$

The integral here is a Stieltjes integral. The η_k's are called *moments.* A sequence $\{\eta_k\}$ which arises in this way is called a *moment sequence* arising from x. A moment sequence is certainly bounded, for $|\eta_k|$ cannot exceed the total variation of x. But which bounded sequences are moment sequences? That is, how can we recognize those bounded sequences which are in the range of A? This is a classical problem known as the *moment problem of Hausdorff.* In this case $\mathscr{R}(A)$ is not all of Y. For a description of $\mathscr{R}(A)$ (i.e., of all moment sequences) we refer the reader to Chapter 3 of D. V. Widder, **1**; consult the bibliography.*

In addition to existence problems there are *uniqueness* problems. In the case of any linear operator A we can ask: is the x such that $Ax = y$ unique in all cases in which a solution for x exists? This is the same as the question: does A^{-1} exist? By Theorem 1.3–B this is equivalent to the question: does the equation $Ax = 0$ have a unique solution (namely $x = 0$)? To answer this question for a particular operator we must have a good deal of detailed knowledge about the operator.

Example 3. Let $Y = C[a, b]$ and let X be the subspace of $C[a, b]$ consisting of those functions x which have continuous first and second derivatives on $[a, b]$ and are such that $x(a) = x(b) = 0$. Define A on X into Y by $Ax = x''$, where x'' is the second derivative of x. It is easy to show that A^{-1} exists. For, $Ax = 0$ implies that $x(s) = c_1 s + c_2$, where c_1 and c_2 are constants, and the requirements $x(a) = x(b) = 0$, then lead to the conclusions $c_1 = c_2 = 0$. It is also easy to show that $\mathscr{R}(A) = Y$. For a given $y \in C[a, b]$ the unique x in X such that $Ax = y$ is given by

$$x(s) = \int_a^s du \int_a^u y(t)\, dt - \frac{s - a}{b - a} \int_a^b du \int_a^u y(t)\, dt.$$

With a little manipulation this formula may be put in the form

$$x(s) = \int_a^b K(s, t) y(t)\, dt,$$

where

$$K(s, t) = \begin{cases} \dfrac{(s - b)(t - a)}{b - a} & \text{if } a \leqslant t \leqslant s, \\[2ex] \dfrac{(s - a)(t - b)}{b - a} & \text{if } s \leqslant t \leqslant b. \end{cases}$$

Note that $K(s, t) = K(t, s)$ and that K is continuous on the square where it is defined.

Sometimes there is an important connection between existence problems and uniqueness problems. This connection is simplest in the case of finite-dimensional spaces, as we shall see in Theorem 1.3–E.

---

* References to the bibliography are made by listing the author's name and a number in boldface type, identifying a book or article by that author.

Lemma 1.3–C. *Let A be a linear operator on X into Y, and let X be of finite dimension n. Then $\mathscr{R}(A)$ is of some finite dimension m, with $m \leqslant n$.*

PROOF. If $\mathscr{R}(A) = (0)$, then $m = 0$, and there is nothing more to prove. Since $X = (0)$ implies $\mathscr{R}(A) = (0)$, we assume that $n > 0$ and that $\mathscr{R}(A) \neq (0)$. Suppose $y_1, \cdots, y_{n+1}$ are in $\mathscr{R}(A)$. Then, there exist vectors $x_1, \cdots, x_{n+1}$ in X with $Ax_k = y_k$, $k = 1, \cdots, n + 1$. Since X is of dimension n, there exists scalars $\alpha_1, \cdots, \alpha_{n+1}$, not all zero, such that $\alpha_1 x_1 + \cdots + \alpha_{n+1}x_{n+1} = 0$. Then $0 = A(\alpha_1 x_1 + \cdots + \alpha_{n+1}x_{n+1}) = \alpha_1 y_1 + \cdots + \alpha_{n+1}y_{n+1}$, so that the set $y_1, \cdots, y_{n+1}$ is linearly dependent. Now $\mathscr{R}(A)$ certainly contains some finite linearly independent sets, for a single nonzero vector is such a set. Hence, $\mathscr{R}(A)$ contains a finite linearly independent set with the largest possible number of elements, and this number m cannot exceed n, as has been shown. This completes the proof.

Theorem 1.3–D. *In addition to the hypothesis of Lemma* 1.3–C, *assume that A^{-1} exists. Then $\mathscr{R}(A)$ is of the same dimension as X (the domain of A).*

PROOF. We know $m \leqslant n$, by the lemma. Now $\mathscr{R}(A) = \mathscr{D}(A^{-1})$ and $X = \mathscr{R}(A^{-1})$. Hence, applying the lemma to A^{-1}, $n \leqslant m$. Thus $m = n$.

Theorem 1.3–E. *Let A be a linear operator on X into Y, where X and Y are both of the same finite dimension n. Then $\mathscr{R}(A) = Y$ if and only if A^{-1} exists.*

PROOF. If A^{-1} exists, we know that $\mathscr{R}(A)$ is of the same dimension as X. It follows by Theorem 1.1–E that $\mathscr{R}(A)$ cannot be a proper subset of Y and must, therefore, coincide with Y. On the other hand, suppose that $\mathscr{R}(A) = Y$. Let $y_1, \cdots, y_n$ be a basis of Y. There exist vectors $x_1, \cdots, x_n$ in X such that $Ax_k = y_k$, $k = 1, 2, \cdots, n$, since $\mathscr{R}(A) = Y$. The set of x's is linearly independent. For, if $\alpha_1 x_1 + \cdots + \alpha_n x_n = 0$, it follows that $0 = A(\alpha_1 x_1 + \cdots + \alpha_n x_n) = \alpha_1 y_1 + \cdots + \alpha_n y_n$, whence $\alpha_1 = \cdots = \alpha_n = 0$, since the set of y's is linearly independent. Consequently, the set $x_1, \cdots, x_n$ is a basis of X. Now suppose $Ax = 0$ for some x. We can express x in the form $x = \alpha_1 x_1 + \cdots + \alpha_n x_n$, and so $0 = Ax = \alpha_1 y_1 + \cdots + \alpha_n y_n$. This implies $\alpha_1 = \cdots = \alpha_n = 0$, and so $x = 0$. Therefore A^{-1} exists. The proof is now complete.

Finite dimensionality is essential in Theorem 1.3–E. We can see that this is so by the following example. Let X be the linear space of Example 9, § 1.2, and let $Y = X$. Define A by setting $Ax = y$, where $y(s) = x'(s)$. In this case $\mathscr{R}(A) = X$, for a solution of $Ax = y$ is given by $x(s) = \int_0^s y(t)\,dt$. The inverse A^{-1} does not exist, however, for $Ax = 0$ in the case of every constant function.

PROBLEM

Alternative argument for second part of proof of Theorem 1.3–E: Assume that $\mathscr{R}(A) = Y$ but that A^{-1} does not exist, so that there is an $x_1 \neq 0$ with $Ax_1 = 0$. Choose $x_2, \cdots, x_n$ so that $x_1, \cdots, x_n$ is a basis of X. Let $y_k = Ax_k$, $k = 2, \cdots, n$, and show that the whole of Y is generated by $y_2, \cdots, y_n$.

1.4 Linear Operators in Finite Dimensional Spaces

Let X denote either the real space R_n or the complex space C_n (Examples 1 and 2, § 1.2). Let $Y = R_m$ if $X = R_n$, and $Y = C_m$ if $X = C_n$. Here m and n may be any positive integers. Let

$$(\alpha_{ij}) \equiv \left\| \begin{matrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{m1} & \alpha_{m2} & \cdots & \alpha_{mn} \end{matrix} \right\| \tag{1.4–1}$$

be any $m \times n$ matrix of scalars (real or complex according as X and Y are real or complex). Such a matrix defines a linear operator A on X into Y as follows: $Ax = y$, where $x = (\xi_1, \cdots, \xi_n)$, $y = (\eta_1, \cdots, \eta_m)$, and

$$\sum_{j=1}^{n} \alpha_{ij}\xi_j = \eta_i \qquad i = 1, \cdots, m. \tag{1.4–2}$$

The linear problems connected with the operator A are, in this case, problems connected with the system of m linear equations (1.4–2). The special properties of the operator A must be found by an examination of the particular matrix (1.4–1).

We recall the notation

$$e_1 = (1, 0, \cdots, 0), \cdots, e_n = (0, \cdots, 0, 1)$$

introduced in Example 1, § 1.2. The vectors $e_1, \cdots, e_n$ form a basis of X. Let us write

$$f_1 = (1, 0, \cdots, 0), \cdots, f_m = (0, \cdots, 0, 1)$$

for the analogous basis in Y. Observe that $x = \xi_1 e_1 + \cdots + \xi_n e_n$ and $y = \eta_1 f_1 + \cdots + \eta_m f_m$. Now let us define vectors $a_1, \cdots, a_n$ in Y by $a_k = Ae_k$. Since $\xi_j = \delta_{kj}$ (the Kronecker delta) if $x = e_k$, we see by (1.4–2) that

$$a_k = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \alpha_{ij}\delta_{kj} \right) f_i = \sum_{i=1}^{m} \alpha_{ik} f_i = (\alpha_{1k}, \alpha_{2k}, \cdots, \alpha_{mk}).$$

In other words, the vector a_k appears as the k^{th} column in the matrix (1.4–1).

In the foregoing situation we started with an arbitrary $m \times n$ matrix and used it to define a linear operator on X into Y. Let us now proceed the other way around. We shall start with an arbitrary linear operator A on X into Y and show that there is an $m \times n$ matrix of the form (1.4–1) such that it defines the operator A, in the sense that $Ax = y$ means precisely that the η's are defined in terms of the ξ's by equations (1.4–2).

If A is the given operator, we define $a_j = Ae_j$, $j = 1, \cdots, n$. The a's are vectors in Y and can, therefore, be expressed in terms of the basis $f_1, \cdots, f_m$. Let the expression of a_j be $a_j = \alpha_{1j}f_1 + \cdots + \alpha_{mj}f_m$. In this way we arrive at the set of scalars α_{ij}, with which we form the matrix (1.4–1). Now, if $x = (\xi_1, \cdots, \xi_n)$ and $y = Ax = (\eta_1, \cdots, \eta_m)$, we have to show that the equations (1.4–2) are valid. We have $x = \xi_1 e_1 + \cdots + \xi_n e_n$, $Ax = \xi_1 a_1 + \cdots + \xi_n a_n$, or

$$y = \sum_{j=1}^{n} \xi_j \left(\sum_{i=1}^{m} \alpha_{ij} f_i \right) = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \alpha_{ij} \xi_j \right) f_i.$$

Since the expression of a vector in terms of the basis is unique, this last equation is equivalent to the equations (1.4–2).

The foregoing considerations show that the study of linear operators on R_n into R_m is closely related to the study of matrices of the form (1.4–1) with real elements α_{ij}. If we deal with C_n and C_m the only difference is that the matrix elements are from the complex number field.

If we now turn to a study of linear operators on X into Y, where X and Y are *arbitrary* finite-dimensional spaces, we shall find the same close connection between linear operators and matrices. The reason for this is that every n-dimensional real linear space X is in a certain sense the same as R_n (or the same as C_n if X is complex instead of real). This sense in which two spaces are said to be the same will now be explained.

Definition. Two linear spaces X, Y (with the same scalar field) are said to be *isomorphic* if there exists a linear operator T whose domain is X, whose range is all of Y, and whose inverse T^{-1} exists.

In less technical language, X and Y are isomorphic if there is a one-to-one correspondence between the elements of X and the elements of Y, such that the operations of vector addition and scalar multiplication are preserved under the correspondence. That is, if x_1 and x_2 have correspondents y_1 and y_2, respectively, then $y_1 + y_2$ corresponds to $x_1 + x_2$ and αy_1 corresponds to αx_1.

If X is isomorphic to Y and Y is isomorphic to Z, it is easy to see that X is isomorphic to Z.

Theorem 1.4–A. *If X and Y are both n-dimensional linear spaces with the same scalar field, they are isomorphic.*

PROOF. Assume the scalar field is that of the real numbers, for definiteness. We shall show that X (and hence also Y) is isomorphic to R_n, from which it will follow that X and Y are isomorphic. Let $x_1, \cdots, x_n$ be a basis of X. Every x in X has a unique representation $x = \xi_1 x_1 + \cdots + \xi_n x_n$, where $\xi_1, \cdots, \xi_n$ are real numbers. We define a linear operator T on X into R_n by writing $Tx = (\xi_1, \cdots, \xi_n)$. The facts that this is a linear operator, that it has all of R_n as its range, and that T^{-1} exists are all easily verified, and we omit the details. This completes the proof.

If X is n-dimensional, with a basis $x_1, \cdots, x_n$, the coefficients $\xi_1, \cdots, \xi_n$ in the representation $x = \xi_1 x_1 + \cdots + \xi_n x_n$ may be called the *coordinates* (or, also, the *components*) of x with respect to the basis $x_1, \cdots, x_n$. Thus, the isomorphism of X and R_n is established by correlating x with the point in R_n whose coordinates with respect to the basis $e_1, \cdots, e_n$ of R_n are the same as the coordinates of x with respect to $x_1, \cdots, x_n$.

Suppose now that X and Y are any linear spaces of dimensions n and m, respectively, with the same scalar field. Let $x_1, \cdots, x_n$ be a basis in X, and $y_1, \cdots, y_m$ a basis in Y. Let A be a linear operator on X into Y. Since $Ax_j \in Y$, we can write

$$(1.4\text{–}3) \qquad Ax_j = \sum_{i=1}^{m} \alpha_{ij} y_i, \qquad j = 1, \cdots, n.$$

The operator A, in conjunction with the two bases $(x_1, \cdots, x_n)$, $(y_1, \cdots, y_m)$, determines an $m \times n$ matrix (α_{ij}). With this matrix we can calculate the vector Ax for every $x \in X$. The calculations are essentially the same as those given earlier in the special case $X = R_n$, $Y = R_m$; the coordinates η_i of Ax are given in terms of the coordinates ξ_j of x by equations (1.4–2).

It must be kept in mind that, although the operator A is represented by the matrix (α_{ij}), *it is not the same thing as the matrix.* For, the operator A is not dependent on the particular bases which are chosen for X and Y, whereas the matrix *does* depend on these bases, and the same operator is represented by different matrices when different bases are chosen.

To conclude this section we give an example.

Example. Let X be the complex linear space consisting of all polynomials in the real variable s of degree not exceeding $n - 1$, with complex coefficients. Let Y be the corresponding space of polynomials with degree not exceeding n. It is readily evident that Z is of dimension n, one possible basis $x_1, \cdots, x_n$ being that defined by

$$x_1(s) = 1,\ x_2(s) = s,\ \cdots,\ x_n(s) = s^{n-1}.$$

Likewise, Y is of dimension $n + 1$. X is a subspace of Y. For a basis in Y we take $y_1 = x_1, \cdots, y_n = x_n,\ y_{n+1}(s) = s^n$.

Now, consider the operator A on X into Y defined by $Ax = y$, where

$$y(s) = \int_0^s x(t)\, dt.$$

We see that

$$Ax_j = \frac{1}{j} y_{j+1}, \qquad j = 1, \cdots, n.$$

The matrix (α_{ij}) in this case is therefore

$$\left\| \begin{matrix} 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \frac{1}{2} & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \ddots & \vdots \\ \vdots & \vdots & & & \ddots & 0 \\ 0 & 0 & \cdots & \cdots & 0 & \frac{1}{n} \end{matrix} \right\|$$

with $n + 1$ rows and n columns.

We observe, incidentally, that the inverse A^{-1} exists. The range $\mathscr{R}(A)$ consists of all polynomials of degree $\leqslant n$ whose constant term is zero. That is, $\mathscr{R}(A)$ is the linear manifold spanned by $y_2, \cdots, y_{n+1}$. With $\mathscr{R}(A)$ as its domain, A^{-1} is defined by $A^{-1}y = x$, where $x(s) = y'(s)$.

PROBLEM

Let X and Y be as in the example, but take $x_k = y_k = P_{k-1}$, $k = 1, \cdots, n$, $y_{n+1} = P_n$, where P_n is the Legendre polynomial of degree n. Here we have bases for X and Y different from the ones previously mentioned. The recursion formula

$$(2k + 1)P_k(s) = P'_{k+1}(s) - P'_{k-1}(s)$$

and the known values of $P_k(0)$

$$\begin{aligned} P_k(0) &= 0 \text{ if } k \text{ is odd} \\ &= (-1)^{k/2} \frac{1\cdot 3\cdot 5 \cdots (k-1)}{2\cdot 4 \cdots k} \text{ if } k \text{ is even} \end{aligned}$$

make it possible to compute the matrix of the operator A relative to these bases. For $n = 6$ the matrix is

$$\left\| \begin{matrix} 0 & \frac{1}{6} & 0 & -\dfrac{1}{2\cdot 4} & 0 & \dfrac{1\cdot 3}{2\cdot 4\cdot 6} \\ 1 & 0 & -\frac{1}{5} & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & -\frac{1}{7} & 0 & 0 \\ 0 & 0 & \frac{1}{5} & 0 & -\frac{1}{9} & 0 \\ 0 & 0 & 0 & \frac{1}{7} & 0 & -\frac{1}{11} \\ 0 & 0 & 0 & 0 & \frac{1}{9} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{11} \end{matrix} \right\|$$

1.5 Other Examples of Linear Operators

In this section we give a number of illustrations of problems in analysis, as formulated in terms of linear operators.

Example 1. Let G be a nonempty, bounded, connected, open set in the Euclidean xy-plane. Let $\bar{G}$ be the closure of G; then the boundary $\bar{G}-G$ is nonempty. Let H be the class of all real-valued functions f of x, y which are defined and continuous on $\bar{G}$ and harmonic in G (i.e., each f must satisfy Laplace's equation in G). Let C denote the class of real-valued functions defined and continuous on $\bar{G}-G$. Clearly H and C are both linear spaces.

Now, if $u \in H$, $u(x, y)$ is defined at every point $(x, y) \in \bar{G}$. If we think only of the values of u at the points of $\bar{G}-G$, we arrive at an element f of C defined by $f(x, y) = u(x, y)$ when $(x, y) \in \bar{G}-G$. We call f the restriction of u to the boundary of G. We define a linear operator A on H into C by writing $Au = f$; the fact that A is linear is evident.

Let us consider existence and uniqueness problems for A (see § 1.3 for a general discussion of these problems). In the present example these are problems of potential theory. The uniqueness problem is completely disposed of by the well-known fact that, if $u \in H$ and $u(x, y) = 0$ on $\bar{G}-G$, then $u(x, y) = 0$ at all points of G (since u must attain both its maximum and minimum values on the boundary of G). In our present notation this means that $u = 0$ if $Au = 0$; consequently A^{-1} exists. For the existence problem we want to know if $\mathscr{R}(A)$ is all of C. This is the famous Dirichlet problem: Is there for each $f \in C$ an element $u \in H$ such that f is the restriction of u to the boundary of G? The answer to this question is not always affirmative; $\mathscr{R}(A)$ may be a proper subspace of C. However, under certain conditions involving the nature of the set G and its boundary, it is known that $\mathscr{R}(A) = C$. In particular, under suitable conditions, $\mathscr{R}(A) = C$ and A^{-1} is defined by an integral, involving the Green's function of G, extended over the boundary of G.

Many boundary-value problems of both ordinary and partial differential equations are susceptible of formulation in terms of linear operators, by procedures analogous to those of Example 1. It is essential that the differential equations be linear and that either the differential equations or the boundary conditions be homogeneous.

Example 2. Let $X = Y = C[a, b]$ (see Example 3, § 1.2). Let $k(s, t)$ be defined for $a \leqslant s \leqslant b$, $a \leqslant t \leqslant b$ and such that for each $x \in X$ the Riemann integral

$$\int_a^b k(s, t)x(t)\,dt \tag{1.5-1}$$

exists and defines a continuous function of s on $[a, b]$. The values of k are to be either real or complex, depending on whether the values of the elements of $C[a, b]$ are taken as real or complex.

The integral (1.5-1) defines a linear operator K on X into X, if we take $Kx = y$ to mean

$$y(s) = \int_a^b k(s, t)x(t)\,dt. \tag{1.5-2}$$

The equation (1.5-2) is called an *integral equation.* It is of the particular sort known as an equation *of Fredholm type, of the first kind.*

Another operator T is obtained by defining $Tx = y$ to mean

$$y(s) = x(s) - \int_a^b k(s, t)x(t)\,dt. \tag{1.5-3}$$

In this case the integral equation is said to be *of Fredholm type, of the second kind.* Equations of this sort are of great importance. There is a well worked-out theory of the existence and uniqueness problems associated with such equations. The applications of this theory play a vital role in the theory of boundary-value problems in differential equations.

A special situation results if we assume that $k(s, t) = 0$ when $t > s$. The integral (1.5-1) then becomes

$$\int_a^s k(s, t)x(t)\,dt.$$

The equations (1.5-2) and (1.5-3), in this modified form, are said to be *of Volterra type.* The theory of Volterra equations of the second kind is particularly simple, much more so than the corresponding theory for equations of the Fredholm type. Later on in this book we shall treat the theory of integral equations of the second kind, using abstract-linear-space methods. For this treatment it is necessary to introduce topological as well as algebraic considerations about linear spaces.

Example 3. Let $X = Y = C[0, \pi]$. In this case we assume specifically that we are dealing with complex-valued continuous functions. If $x \in X$ and if $x'(s)$ exists on $[0, \pi]$, we denote the derived function by x'. Likewise x'' denotes the second derivative if $x''(s)$ exists on $[0, \pi]$. We now define a linear operator A as follows: Let $\mathscr{D}(A) = \{x : x', x'' \in C[0, \pi] \text{ and } x(0) = x(\pi) = 0\}$. (See the statement in the Introduction about defining subsets of X by the brace notation.) For $x \in \mathscr{D}(A)$ define $Ax = y$ to mean

$$-x''(s) + \lambda x(s) = y(s), \tag{1.5–4}$$

where λ is a complex parameter. The operator A thus depends on λ. Clearly $\mathscr{R}(A) \subset X$.

We shall discuss the existence and uniqueness problem for A in some detail. This amounts to a discussion of finding solutions of the differential equation (1.5–4) which satisfy the end conditions $x(0) = x(\pi) = 0$.

Let one of the square roots of λ be denoted by μ. By the method of variation of parameters the general solution of (1.5–4) (disregarding the end conditions) is found to be

$$x(s) = \frac{1}{\mu} \int_0^s y(t) \sinh \mu(t - s)\, dt + C_1 e^{\mu s} + C_2 e^{-\mu s}, \tag{1.5–5}$$

where C_1 and C_2 are arbitrary constants. This is in case $\lambda \neq 0$. If $\lambda = 0$ the solution may be found by direct integration. If we impose the end conditions $x(0) = x(\pi) = 0$, we find, first, that $C_1 + C_2 = 0$ and then that

$$0 = \frac{1}{\mu} \int_0^\pi y(t) \sinh \mu(t - \pi)\, dt + 2C_1 \sinh \pi\mu.$$

To solve this equation for C_1 we must assume that $\sinh \pi\mu \neq 0$. On substituting into (1.5–5) the values found for C_1, C_2, we have

$$x(s) = \frac{1}{\mu} \int_0^s y(t) \sinh \mu(t - s)\, dt - \frac{\sinh \mu s}{\mu \sinh \mu\pi} \int_0^\pi y(t) \sinh \mu(t - \pi)\, dt. \tag{1.5–6}$$

The corresponding formula when $\lambda = 0$ is

$$x(s) = \int_0^s y(t)(t - s)\, dt - \frac{s}{\pi} \int_0^\pi y(t)(t - \pi)\, dt. \tag{1.5–7}$$

It may now be seen that the equation $Ax = y$ admits a solution $x \in \mathscr{D}(A)$ for every $y \in X$, provided either that $\lambda = 0$ or that $\sinh \pi\mu \neq 0$

if $\lambda \neq 0$. Moreover, the solution is unique, for it may be argued from the standard theory of linear differential equations that (1.5–6) and (1.5–7) give, in their respective cases, the *only* solutions of (1.5–4) satisfying the given end conditions.

There remain to be considered the cases when $\lambda \neq 0$ but $\sinh \pi\mu = 0$. Now, since $\mu^2 = \lambda$ and since $\sinh \pi\mu = 0$ if and only if $\pi\mu = in\pi$, where n is an integer, we see that the exceptional cases are given by $\lambda = -n^2$, $n = 1, 2, 3, \cdots$. In these cases (1.5–4) cannot have a unique solution in $\mathscr{D}(A)$ (i.e., A^{-1} does not exist), for $Ax = 0$ in case $\lambda = -n^2$ and $x(s) = \sin ns$. If we re-examine our earlier work, assuming that $\lambda = -n^2 \neq 0$, we see that the equation $Ax = y$ has all its solutions $x \in \mathscr{D}(A)$ given by

$$x(s) = \frac{1}{n}\int_0^s y(t) \sin n(t - s)\, dt + C \sin ns, \tag{1.5–8}$$

where C is an arbitrary constant, if and only if the function y satisfies the condition

$$\int_0^\pi y(t) \sin nt\, dt = 0. \tag{1.5–9}$$

This condition serves as a description of $\mathscr{R}(A)$ in this case.

By writing $\int_0^\pi = \int_0^s + \int_s^\pi$ in (1.5–6), and rearranging somewhat, the formula can be written

$$x(s) = \int_0^s y(t) \frac{\sinh \mu t \sinh \mu(\pi - s)}{\mu \sinh \mu\pi}\, dt + \int_s^\pi y(t) \frac{\sinh \mu s \sinh \mu(\pi - t)}{\mu \sinh \mu\pi}\, dt.$$

If we define

$$k(s, t) = \frac{\sinh \mu t \sinh \mu(\pi - s)}{\mu \sinh \mu\pi} \qquad \text{when } 0 \leqslant t \leqslant s$$

and $k(s, t) = k(t, s)$ when $s \leqslant t \leqslant \pi$, (1.5–6) becomes

$$x(s) = \int_s^\pi k(s, t)y(t)\, dt.$$

We note that this formula is of the same type as (1.5–2). In our operator notation this formula is $x = A^{-1}y$. If we regard $x = A^{-1}y$ as an equation to be solved for y, we see that it is an integral equation of Fredholm type, of the first kind. It has a unique solution, for each $x \in \mathscr{D}(A)$, given by (1.5–4).

Example 4. Let $X = Y = L(0, \infty)$ (see Example 7, § 1.2). Let us define an operator A as follows: Consider those functions in $\mathscr{L}(0, \infty)$ which are absolutely continuous on $[0, a]$ for every finite a, with derivatives also belonging to $\mathscr{L}(0, \infty)$. Each such function determines an equivalence class in $\mathscr{L}(0, \infty)$, i.e., an element of $L(0, \infty)$. The set of such elements of $L(0, \infty)$ is going to be $\mathscr{D}(A)$. For convenience we shall drop the notational distinction between an element of $L(0, \infty)$ and one of its representatives in $\mathscr{L}(0, \infty)$. Thus if $x \in L(0, \infty)$, we shall write $x(s)$ to indicate a functional value at s of a representative of x. We then define $Ax = y$ to mean that

$$(1.5\text{–}10) \qquad -x'(s) + \lambda x(s) = y(s)$$

almost everywhere on $(0, \infty)$. This defines $y \in L(0, \infty)$ uniquely if $x \in \mathscr{D}(A)$. As in Example 3, λ denotes a complex parameter.

To solve the equation $Ax = y$, we proceed formally to solve (1.5–10) by use of the integrating factor $e^{-\lambda s}$. If $x(s)$ is absolutely continuous and satisfies (1.5–10), it is necessarily of the form

$$(1.5\text{–}11) \qquad x(s) = e^{\lambda s}\left\{C - \int_0^s y(t)e^{-\lambda t}\,dt\right\},$$

where C is some constant, and, vice versa, the function $x(s)$ so defined, for any C, is absolutely continuous and does satisfy (1.5–10). If $x(s)$ turns out to belong to $\mathscr{L}(0, \infty)$, it is clear from (1.5–10) that $x'(s)$ will also belong, because we assume that $y(s)$ belongs to $\mathscr{L}(0, \infty)$. The problem is then that of seeing whether the constant C can be chosen so that the integral $\int_0^\infty |x(s)|\,ds$ will be convergent. In considering this problem it is convenient to consider three cases, depending upon whether the real part of λ is positive, negative, or zero. We write $\lambda = \alpha + i\beta$, α and β real. Note that

$$|x(s)| = e^{\alpha s}\left|C - \int_0^s y(t)e^{-\lambda t}\,dt\right|.$$

Case I. $\alpha > 0$. Since

$$\int_0^s y(t)e^{-\lambda t}\,dt \to \int_0^\infty y(t)e^{-\lambda t}\,dt \qquad \text{as } s \to \infty,$$

we see that $|x(s)|$ will be of the order of magnitude of $e^{\alpha s}$ when s is large unless

$$(1.5\text{–}12) \qquad C = \int_0^\infty y(t)e^{-\lambda t}\,dt.$$

This is therefore a necessary condition for $x(s)$ to belong to $\mathscr{L}(0, \infty)$, when $\alpha > 0$. It is also sufficient; for, if it is satisfied, we have

$$(1.5\text{–}13) \qquad x(s) = e^{\lambda s}\int_s^\infty y(t)e^{-\lambda t}\,dt = \int_0^\infty y(s + u)e^{-\lambda u}\,du,$$

from which it is easy to see that $x(s) \in \mathscr{L}(0, \infty)$ if $\alpha > 0$.

Case 2. $\alpha = 0$. The condition (1.5–12) is likewise necessary in this case. It is not sufficient, however, and whether or not $x(s) \in \mathscr{L}(0, \infty)$ will depend on the choice of y. For instance, if

$$y(t) = e^{i\beta t}t^{-2}$$

when t is large, then for large s

$$x(s) = e^{\lambda s}\int_s^\infty y(t)e^{-\lambda t}\,dt = e^{i\beta s}\int_s^\infty t^{-2}\,dt = \frac{e^{i\beta s}}{s},$$

and $x(s)$ is not in $\mathscr{L}(0, \infty)$.

Case 3. $\alpha < 0$. In this case the function $x(s)$ defined by (1.5–11) is in $\mathscr{L}(0, \infty)$, no matter how C and y are chosen. This is easily seen, for $e^{\lambda s}$ is in $\mathscr{L}(0, \infty)$, and

$$\int_0^s y(t)e^{\lambda(s-t)}\,dt$$

is also in $\mathscr{L}(0, \infty)$. In fact,

$$\int_0^\infty \left|\int_0^s y(t)e^{\lambda(s-t)}\,dt\right| ds \leqslant \int_0^\infty ds \int_0^s |y(t)|e^{\alpha(s-t)}\,dt$$
$$= \int_0^\infty e^{-\alpha t}|y(t)|\,dt\int_t^\infty e^{\alpha s}\,ds = -\frac{1}{\alpha}\int_0^\infty |y(t)|\,dt.$$

We can sum up the discussion as follows. The inverse A^{-1} exists if $\alpha \geqslant 0$, and the solution $x = A^{-1}y$ is expressed by (1.5–13). If $\alpha > 0$, $\mathscr{R}(A)$ is all of X, but not so if $\alpha = 0$. When $\alpha < 0$, A^{-1} does not exist, but $\mathscr{R}(A)$ is all of X.

Example 5. Let $X = L(0, 2\pi)$. Let F be the class of functions f analytic in the unit circle $|z| < 1$. For $x \in X$, with representative $x(s) \in \mathscr{L}(0, 2\pi)$, define $Ax = f$ to mean

$$f(z) = \frac{1}{2\pi}\int_0^{2\pi} \frac{x(s)}{1 - ze^{-is}}\,ds. \tag{1.5–14}$$

Then A is a linear operator on X into F. The inverse A^{-1} does not exist, for it is a simple matter to calculate that $Ax = 0$ if $x(s) = e^{-ims}$, $m = 1, 2, \cdots$ (use the geometric series for $(1 - ze^{-is})^{-1}$, and integrate term by term). The problem of characterizing the range $\mathscr{R}(A)$ has not yet been solved satisfactorily. One would like to know conditions on the analytic function f, necessary and sufficient in order that f can be represented by formula (1.5–14), with some $x \in L(0, \pi)$. A simple *sufficient*, but not necessary, condition is that f be continuous when $|z| \leqslant 1$ as well as analytic when $|z| < 1$. For then, by Cauchy's formula,

$$f(z) = \frac{1}{2\pi i}\int \frac{f(w)}{w - z}\, dw,$$

the complex line integral being extended counterclockwise around the unit circle $|w| = 1$. If we write $w = e^{is}$, $0 \leqslant s \leqslant 2\pi$, this becomes

$$f(z) = \frac{1}{2\pi}\int_0^{2\pi} \frac{f(e^{is})}{1 - ze^{-is}}\, ds.$$

Thus in this case $f = Ax$, where $x(s) = f(e^{is})$.

A less restrictive sufficient condition is known. Suppose $f \in F$, and suppose that there is a constant M such that

$$\int_0^{2\pi} |f(re^{i\theta})|\, d\theta \leqslant M \tag{1.5–15}$$

when $0 \leqslant r < 1$. Then it may be shown that $\lim_{r\to 1} f(re^{is}) = x(s)$ exists almost everywhere on $0 \leqslant s \leqslant 2\pi$ and that $x(s) \in \mathscr{L}(0, 2\pi)$. It may also be shown that

$$\lim_{r\to 1}\int_0^{2\pi} |f(re^{i\theta}) - x(\theta)|\, d\theta = 0. \tag{1.5–16}$$

See § 7.56 of Zygmund, **1**; these results are due to F. Riesz, **2**. From these facts it is easy to prove that $Ax = f$. Thus, a sufficient condition that $f \in \mathscr{R}(A)$ is that f be an element of F satisfying (1.5–15) for some constant M and all r such that $0 \leqslant r < 1$.

Example 6. Let X be the class of all infinite arrays of the form $x = (\cdots, \xi_{-2}, \xi_{-1}, \xi_0, \xi_1, \xi_2, \cdots)$, where the ξ_n's are arbitrary complex numbers. We denote these more briefly as $x = \{\xi_n\}$, $n = 0, \pm 1, \pm 2, \cdots$. Define addition and scalar multiplication in the obvious way, i.e.,

$$\{\xi_n\} + \{\eta_n\} = \{\xi_n + \eta_n\}, \quad \alpha\{\xi_n\} = \{\alpha\xi_n\}.$$

Then X is a complex linear space. Now consider the linear operator A on $L(0, \pi)$ into X defined as follows: If $f \in L(0, 2\pi)$ with representative $f(t) \in \mathscr{L}(0, 2\pi)$, let $Af = x$, where $x = \{\xi_n\}$ is defined by

$$\xi_n = \frac{1}{2\pi}\int_0^{2\pi} f(t)e^{-int}\, dt, \qquad n = 0, \pm 1, \pm 2, \cdots.$$

The numbers ξ_n are the complex Fourier coefficients of f.

The inverse A^{-1} exists. This is simply the classical theorem that if $f(t) \in \mathscr{L}(0, 2\pi)$ and if all of its Fourier coefficients vanish, then $f(s) = 0$

almost everywhere [so that $f = 0$ as an element of $L(0, 2\pi)$]. The problem of characterizing $\mathscr{R}(A)$ has never been fully solved, though many things are known about $\mathscr{R}(A)$ from the theory of Fourier series. Many interesting and important results have been found by confining attention to subspaces of $L(0, 2\pi)$ and X. For instance, it is known that if we consider the domain of A to be $L^2(0, 2\pi)$ instead of $L(0, 2\pi)$, the range of A is the subspace of X consisting of all $\{\xi_n\}$ for which the series

$$\sum_{n=-\infty}^{\infty} |\xi_n|^2$$

is convergent.

PROBLEM

Suggestion for showing that $Ax = f$ in the context of the paragraph where (1.5–15) occurs: Let $C(z, \theta) = (1 - ze^{-i\theta})^{-1}$. If $|z| < r < 1$, it may be seen from Cauchy's integral formula that

$$\left| f(z) - \frac{1}{2\pi} \int_0^{2\pi} C(z, \theta) x(\theta)\, d\theta \right|$$
$$\leqslant \frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})| \left| C\left(\frac{z}{r}, \theta\right) - C(z, \theta) \right| d\theta + \frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta}) - x(\theta)|\, |C(z, \theta)|\, d\theta.$$

One may then use (1.5–15) and (1.5–16) to obtain $Ax = f$.

1.6 Linear Functionals

The field of real numbers is a real linear space; it is 1-dimensional. Likewise, the field of complex numbers is a 1-dimensional complex linear space.

Let X be a real or complex linear space, and let Y be the linear space composed by the scalar field associated with X. A linear operator on X into Y is then called a *linear functional* on X.

Since we shall have many occasions to deal with linear functionals, we shall make certain conventions of notation concerning them. If X is a linear space and x is a generic symbol for elements of X, we shall often use x' as a generic symbol for linear functionals on X. If $x_1 \in X$ and if x_1' is a linear functional on X, there is no implication of a special relationship between x_1 and x_1'. Since x' is a function, we can speak of its value at the point x; by our previous usage this value would be denoted by

$x'(x)$, or simply $x'x$. It will sometimes be convenient to have another notation. We shall write

$$\langle x, x' \rangle$$

for the value of x' at x. This notation has several advantages, as we shall see, especially when we replace x' by a more complicated symbol representing a linear functional on X.

If x_1' and x_2' are linear functionals on X and α in a scalar, we define $x_1' + x_2'$ and $\alpha x_1'$ by the formulas

$$\langle x, x_1' + x_2' \rangle = \langle x, x_1' \rangle + \langle x, x_2' \rangle, \tag{1.6–1}$$

$$\langle x, \alpha x_1' \rangle = \alpha \langle x, x_1' \rangle. \tag{1.6–2}$$

These are the usual definitions for adding functions and multiplying them by scalars. The functional x' such that $x'(x) = 0$ for every x is denoted by 0. Thus, we use 0 to denote the zero scalar, the zero vector in X, and the zero functional; the meaning of 0 in a particular occurrence will be clear from the context.

In addition to the formulas (1.6–1) and (1.6–2), we note the formulas

$$\langle x_1 + x_2, x' \rangle = \langle x_1, x' \rangle + \langle x_2, x' \rangle, \tag{1.6–3}$$

$$\langle \alpha x, x' \rangle = \alpha \langle x, x' \rangle. \tag{1.6–4}$$

These are merely the properties (1.3–1), (1.3–2) of linear operators, expressed in the new notation for linear functionals.

Definition. Let X be a linear space. Let X^f be the class of all linear functionals on X. By the foregoing definitions X^f becomes a linear space. We call it the *algebraic conjugate* of X.

We shall see later that the space X^f plays an important role in the study of linear operators with domain X.

Before going further we pause to give some illustrative examples of linear functionals. We refer to the examples of linear spaces in § 1.2.

Example 1. Let X be either R_n or C_n, where $n \geqslant 1$ (see Examples 1, 2, of § 1.2). Since a linear functional on X is a linear operator with its range in the 1-dimensional space R_1 or C_1 (depending upon whether X is R_n or C_n), we can see from § 1.4 that every linear functional $x' \in X^f$ has a representation

$$\langle x, x' \rangle = \alpha_1 \xi_1 + \cdots + \alpha_n \xi_n, \qquad x = (\xi_1, \cdots, \xi_n), \tag{1.6–5}$$

where $\alpha_1, \cdots, \alpha_n$ are scalars, and conversely, that every n-tuple of scalars

$(\alpha_1, \cdots, \alpha_n)$ can be used to define a functional x' by formula (1.6–5). From this we see that, if we define

$$Ax' = (\alpha_1, \cdots, \alpha_n),$$

then A is a linear operator on X^f into X. We also see that A^{-1} exists and that $\mathscr{R}(A) = X$. Hence, by the definition in § 1.4, X^f and X are isomorphic in case $X = R_n$ or $X = C_n$.

Example 2. Let $X = C[a, b]$ (see Example 3, § 1.2). Let s_0 be a fixed point in the interval $[a, b]$. The formula

$$\langle x, x' \rangle = x(s_0) \tag{1.6–6}$$

defines an element of X^f. Another example of a linear functional is furnished by

$$\langle x, x' \rangle = \int_a^b x(s)\, df(s), \tag{1.6–7}$$

where f is an arbitrary function of bounded variation defined on $[a, b]$ and the integral is a Stieltjes integral. In case f has a continuous derivative, (1.6–7) can be written

$$\langle x, x' \rangle = \int_a^b x(s) f'(s)\, ds,$$

the integral being a Riemann integral.

It may be observed that (1.6–6) is a special case of (1.6–7). If $s_0 = a$, (1.6–7) reduces to (1.6–6) when we define $f(a) = 0$ and $f(s) = 1$ if $a < s \leqslant b$. If $a < s_0 \leqslant b$, (1.6–7) reduces to (1.6–6) by choosing $f(s) = 0$ when $a \leqslant s < s_0$ and $f(s) = 1$ when $s_0 \leqslant s \leqslant b$.

Example 3. Let $X = l^2$ (see Example 6, § 1.2). Suppose that $a = \{\alpha_n\}$ is an element of l^2, and define $x' \in X^f$ by

$$\langle x, x' \rangle = \sum_{i=1}^{\infty} \alpha_i \xi_i, \qquad \text{where } x = \{\xi_n\}. \tag{1.6–8}$$

The series does converge, for, by Cauchy's inequality,

$$\sum_{i=1}^{n} |\alpha_i \xi_i| \leqslant \left(\sum_{i=1}^{n} |\alpha_i|^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^{n} |\xi_i|^2\right)^{\frac{1}{2}}$$

for every value of n. Therefore, since $a \in l^2$ and $x \in l^2$, the partial sums of the series $\sum |\alpha_i \xi_i|$ are bounded, and the series in (1.6–8) is absolutely convergent. It is evident that x', so defined, is a member of X^f.

Example 4. Let $X = L(-\infty, \infty)$ (see Example 7, § 1.2). If f is a bounded measurable function defined almost everywhere on $(-\infty, \infty)$, we can use it to define a linear functional on X as follows:

$$\langle x, x' \rangle = \int_{-\infty}^{\infty} x(s)f(s)\,ds, \tag{1.6–9}$$

where $x(s)$ is a function in $\mathscr{L}(-\infty, \infty)$ representing the element $x \in L(-\infty, \infty)$, and the integral is a Lebesgue integral.

The reader will note that it was only in the case of Example 1 that we asserted anything about giving a general representation of *all* elements of X^f. Ordinarily, in dealing with the infinite-dimensional spaces arising in analysis, we find it most useful to limit attention to certain subspaces of X^f. These subspaces are arrived at by imposing a topological structure on the space X and then considering only those linear functionals on X which are continuous with respect to the topology in question. We shall see later that the functionals exhibited in Examples 2, 3, and 4 are typical for the respective spaces when suitable topologies are introduced.

Since X^f is a linear space, we may also consider its algebraic conjugate, which we denote by

$$(X^f)^f, \qquad \text{or simply } X^{ff}.$$

We shall denote elements of X^{ff} generically by x'', and we shall use the notations

$$x''(x') = \langle x', x'' \rangle$$

for the value of x'' at x'.

Corresponding to each x in X there is a unique x'' in X^{ff} defined by

$$\langle x', x'' \rangle = \langle x, x' \rangle \tag{1.6–10}$$

or equivalently, $x''(x') = x'(x)$. That this formula does in fact define a linear functional on X^f is apparent from (1.6–1) and (1.6–2). The correspondence between x and x'' in (1.6–10) defines a linear operator J on X into X^{ff}: $Jx = x''$. The fact that J is linear is apparent from (1.6–3) and (1.6–4).

Definition. The mapping of X in X^{ff} by the operator J is called the *canonical mapping* of X into X^{ff}.

It will be shown in Theorem 1.71–C that J has an inverse, that is, that $Jx = 0$ implies $x = 0$. If X is finite dimensional, this is shown in Theorem 1.61–B. It follows from this fact that the range of J is isomorphic to X. Hence we may identify X and the range of J, thus permitting ourselves to look upon X as a subspace of X^{ff}. When we adopt this point of view, we speak of the canonical imbedding of X in X^{ff}.

Definition. If $\mathscr{R}(J) = X^{ff}$, we say that X is *algebraically reflexive.*

It turns out, as we shall see later, that X is algebraically reflexive if and only if it is finite dimensional (Theorem 1.72–A).

PROBLEM

Show that, if X is algebraically reflexive, so is X^f. Suggestion: Let J_0 and J_1 be the canonical mappings of X and X^f into X^{ff} and X^{fff} respectively. The problem is to show that $\mathscr{R}(J_1) = X^{fff}$. If $x''' \in X^{fff}$ is given, define $x' \in X^f$ by the formula $\langle x, x' \rangle = \langle J_0 x, x''' \rangle$. Show that $J_1 x' = x'''$.

1.61. Linear Functionals in Finite-Dimensional Spaces

Let X be an n-dimensional linear space, and let $x_1, \cdots, x_n$ be a basis of X. Let $x = \xi_1 x_1 + \cdots + \xi_n x_n$ be the representation of an element $x \in X$. For any $x' \in X^f$ we have

$$x'(x) = \xi_1 x'(x_1) + \cdots + \xi_n x'(x_n). \tag{1.61–1}$$

The coefficients $x'(x_1), \cdots, x'(x_n)$ are independent of x. This set of coefficients may be prescribed arbitrarily; that is, given any n scalars $\alpha_1, \cdots, \alpha_n$, there exists a unique $x' \in X^f$ such that $x'(x_i) = \alpha_i$, $i = 1, \cdots, n$. This x' is given by

$$x'(x) = \sum_{i=1}^{n} \alpha_i \xi_i, \qquad \text{where } x = \sum_{i=1}^{n} \xi_i x_i. \tag{1.61–2}$$

The case in which one α_i is 1 and the others are 0 is of especial interest. If $\alpha_1 = 1, \alpha_2 = \cdots = \alpha_n = 0$, denote the corresponding x' by x_1'. Then $x_1'(x_1) = 1$, $x_1'(x_2) = \cdots = x_1'(x_n) = 0$. We define $x_2', \cdots, x_n'$ in similar fashion. Specifically, for each $k (k = 1, \cdots, n)$ let $\alpha_i = \delta_{ik}$, and denote the corresponding x' in (1.6–2) by x_k'. The essential character of the functionals $x_1', \cdots, x_n'$ is then exhibited in the equations

$$\langle x_j, x_k' \rangle = \delta_{jk} \qquad j, k = 1, \cdots, n. \tag{1.61–3}$$

To demonstrate this we observe that

$$x_j = \sum_{i=1}^{n} \delta_{ij} x_i, \quad x_k'(x_j) = \sum_{i=1}^{n} \delta_{ik}\, \delta_{ij} = \delta_{jk}.$$

Theorem 1.61–A. *The set $x_1', \cdots, x_n'$ defined in the foregoing discussion is a basis of X^f. Hence the algebraic conjugate of an n-dimensional linear space is n-dimensional.*

PROOF. The set $x_1', \cdots, x_n'$ is linearly independent. For, from $\beta_1 x_1' + \cdots + \beta_n x_n' = 0$ it follows with the aid of (1.61–3) that

$$0 = \left\langle x_j, \sum_{i=1}^{n} \beta_i x_i' \right\rangle = \sum_{i=1}^{n} \beta_i \langle x_j, x_i' \rangle = \sum_{i=1}^{n} \beta_i \, \delta_{ji} = \beta_j,$$

or $\beta_1 = \cdots = \beta_n = 0$. Next, for a given $x' \in X^f$ we define $\alpha_i = x'(x_i)$ and prove that

$$x' = \alpha_1 x_1' + \cdots + \alpha_n x_n', \tag{1.61–4}$$

thereby completing the proof of the theorem. The meaning of (1.61–4) is that, for every $x \in X$,

$$x'(x) = \alpha_1 x_1'(x) + \cdots + \alpha_n x_n'(x). \tag{1.61–5}$$

The validity of this formula is readily verified with the aid of (1.61–2) and (1.61–3).

Definition. The basis $x_1', \cdots, x_n'$ of X^f, as defined in the foregoing work, is said to be *dual* to the basis $x_1, \cdots, x_n$ of X.

Corresponding to each basis of X there is a uniquely defined dual basis of X^f.

Theorem 1.61–B. *If X is n-dimensional and if $x \in X$ is such that $x'(x) = 0$ for every $x' \in X^f$, it follows that $x = 0$. In other words, $Jx = 0$ implies $x = 0$, so that J^{-1} exists.*

PROOF. We use the basis $x_1, \cdots, x_n$ and its dual $x_1', \cdots, x_n'$. In the notation of (1.61–4) and (1.61–2) we have $\alpha_1\xi_1 + \cdots + \alpha_n\xi_n = 0$ for all choices of $\alpha_1, \cdots, \alpha_n$. This implies $\xi_1 = \cdots = \xi_n = 0$, and so $x = 0$.

Theorem 1.61–C. *If X is an n-dimensional linear space, it is algebraically reflexive.*

PROOF. We assume $n > 0$, since the case $n = 0$ is trivial. By Theorem 1.61–A both X^f and X^{ff} are n-dimensional. Consider the canonical mapping J of X into X^{ff} (see § 1.6). Since J^{-1} exists, $\mathscr{R}(J)$ is n-dimensional by Theorem 1.3–D. Hence $\mathscr{R}(J) = X^{ff}$, by Theorem 1.1–E. This completes the proof.

Theorem 1.61–D. *Let M be a proper subspace of the n-dimensional linear space X, and suppose $x_0 \in X - M$. Then there exists an element $x_0' \in X^f$ such that $x_0'(x) = 0$ if $x \in M$ and $x_0'(x_0) = 1$.*

PROOF. Suppose M is m-dimensional. The trivial case $m = 0$ is left to the reader, and we consider the case $0 < m < n$. Let $x_1, \cdots, x_m$ be a basis of M, and write $x_{m+1} = x_0$. The set $x_1, \cdots, x_{m+1}$ is linearly independent. By Theorem 1.1–D we can find a basis of X whose first $m + 1$ vectors are $x_1, \cdots, x_{m+1}$. Let $x_1', \cdots, x_n'$ be the dual basis of X^f, and define $x_0' = x'_{m+1}$. Then the conditions required in the theorem are satisfied.

The proofs of Theorems 1.61–B and 1.61–D have been given by methods depending on finite dimensionality. But the theorems are true if the assumption of finite dimensionality of X is dropped. The proofs in this general case require more profound methods, however; see § 1.71.

PROBLEM

If X^f is n-dimensional, so is X.

1.7 Zorn's Lemma

What is commonly known as Zorn's lemma was originally formulated by Zorn as a proposition in the theory of sets, which he showed to be equivalent to the axiom of choice. Subsequently, this proposition has been given a formulation in terms of partially ordered sets which is extremely convenient for many applications. In this book we state Zorn's lemma and use it as the need arises. It is immaterial whether the reader chooses to regard the lemma as an axiom of set theory or as a theorem derived from the axiom of choice.

Our statement of Zorn's lemma depends upon the notion of a partially ordered set.

Definition. Let P be a set of elements. Suppose there is a binary relation defined between certain pairs of elements a, b of P, expressed symbolically by $a \prec b$, with the properties:

1. If $a \prec b$ and $b \prec c$, then $a \prec c$.
2. If $a \in P$ then $a \prec a$.
3. If $a \prec b$ and $b \prec a$, then $a = b$.

Then P is said to be *partially ordered* by the relation.

For example, if P is the set of all subsets of a given set X, the set inclusion relation ($A \subset B$) gives a partial ordering of P. The set of all complex numbers $z = x + iy$, $w = u + iv$, $\cdots$ is partially ordered by defining $z \prec w$ to mean that $x \leqslant u$ and $y \leqslant v$, where $\leqslant$ has its usual meaning for real numbers.

If P is partially ordered and if, moreover, for every pair a, b in P either $a \prec b$ or $b \prec a$, then P is said to be *completely* ordered. (The adjectives *linearly*, *totally*, and *simply* are also used in place of *completely*.)

For example, the real numbers are completely ordered by the relation "a is less than or equal to b."

A subset of a partially ordered set P is itself partially ordered by the relation which partially orders P. Also, a subset of P may turn out to be completely ordered by this relation.

If P is a partially ordered set and S is a subset of P, an element $m \in P$ is called an *upper bound* of S if $a \prec m$ for every $a \in S$. An element $m \in P$ is said to be *maximal* if $a \in P$ and $m \prec a$ together imply $m = a$.

Zorn's Lemma. *Let P be a nonempty partially ordered set with the property that every completely ordered subset of P has an upper bound in P. Then P contains at least one maximal element.*

Some applications of this proposition will be found in §§ 1.71, 1.72.

1.71 Extension Theorems for Linear Operators

If X is a set, M a proper subset of X, and f a function defined on M, a function F defined on X is called an *extension* of f if $F(x) = f(x)$ when $x \in M$. In practice, we are usually concerned with extensions which conform to certain additional requirements. Often these requirements are of the sort which state that the extension F is to have certain of the properties possessed by f (e.g., be bounded, or continuous, or differentiable, or analytic, etc.). The question as to whether an extension of the required sort exists may be easy or difficult to answer, depending on the particular nature of the problem. In this section we are concerned with the existence of an extension when X is a linear space, M is a proper subspace, and f is a linear operator on M into a second linear space Y. We require the extension F to be linear on X into Y. There may also be further conditions on f and F.

Theorem 1.71–A. *Let X and Y be linear spaces and M a proper subspace of X. Let f be a linear operator, defined on M into Y. Then, there exists a linear operator F defined on X into Y such that F is an extension of f.*

PROOF. Choose any element $x_0 \in X - M$, and let M_0 be the subspace of X spanned by the set consisting of M and x_0. Any element of M_0 can be written in one and only one way in the form $x + \alpha x_0$, where $x \in M$

and α is some scalar. The uniqueness of the representation follows from the fact that, if $x_1 + \alpha_1 x_0 = x_2 + \alpha_2 x_0$, where x_1, $x_2 \in M$, then $x_1 - x_2 = (\alpha_2 - \alpha_1)x_0$, whence $(\alpha_2 - \alpha_1)x_0 \in M$, since M is a linear manifold. But, since $x_0 \notin M$, we are forced to conclude that $\alpha_1 - \alpha_2 = 0$, since otherwise $x_0 = (\alpha_2 - \alpha_1)^{-1}(\alpha_2 - \alpha_1)x_0$ would also be in M. Hence also $x_1 - x_2 = 0$, and the representation is unique.

We now define a linear operator F_0 on M_0 into Y as follows:

$$F_0(x + \alpha x_0) = f(x) + \alpha y_0,$$

where y_0 is some fixed vector in Y. For our present purpose it does not matter how y_0 is chosen. It is readily seen that F_0 is an extension of f and that it is a linear operator on M_0 into Y.

It now seems plausible to suppose that by a continuation of the procedure we can arrive at the required extension F defined on all of X. Indeed, if X is finite dimensional, we shall arrive at the desired goal in a finite number of steps. For the general case the argument is made precise by use of Zorn's lemma.

Let g be a linear operator with domain $\mathscr{D}(g) \subset X$ and range $\mathscr{R}(g) \subset Y$; suppose M is a proper subspace of $\mathscr{D}(g)$ and that g is an extension of f. Let P be the class of all such operators g. If g, $h \in P$, let us define the relation $g \prec h$ to mean that $\mathscr{D}(g) \subset \mathscr{D}(h)$ and that h is an extension of g. This relation defines a partial ordering of P. Moreover, P is nonempty, for certainly $F_0 \in P$.

Now suppose that Q is a completely ordered subset of P. We shall define an element $G \in P$ which is an upper bound of Q. Let $\mathscr{D}(G)$ be the union of all the sets $\mathscr{D}(g)$ corresponding to elements $g \in Q$. This set $\mathscr{D}(G)$ is a subspace of X. For, suppose x_1, $x_2 \in \mathscr{D}(G)$. Then, there exist elements g_1, $g_2 \in Q$ such that $x_k \in \mathscr{D}(g_k)$ $(k = 1, 2)$. We may suppose $g_1 \prec g_2$, since Q is completely ordered. Then $\mathscr{D}(g_1) \subset \mathscr{D}(g_2)$, and so $x_1 + x_2 \in \mathscr{D}(g_2) \subset \mathscr{D}(G)$. Closure of $\mathscr{D}(G)$ under multiplication by scalars is proved even more simply. Now, suppose $x \in \mathscr{D}(G)$. Then $x \in \mathscr{D}(g)$ for some $g \in Q$. We shall define $G(x) = g(x)$. This definition is unambiguous, for, if $x \in \mathscr{D}(g_1)$ and $x \in \mathscr{D}(g_2)$, where g_1, $g_2 \in Q$, we have $g_1(x) = g_2(x)$ by the fact that Q is completely ordered. The proof that G is linear is like the proof that $\mathscr{D}(G)$ is a linear manifold. It is clear that $G \in P$ and that $g \prec G$ for every $g \in Q$.

We now know that P satisfies the conditions of Zorn's lemma and must, therefore, contain a maximal element, say F. The domain of F must be all of X, for otherwise we could regard $\mathscr{D}(F)$ as the M in the first part of the proof and thus obtain an element $g \in P$ with $g \neq F$, $F \prec g$, contrary to the maximality of F. The proof of the theorem is now complete, for F has the properties required in the theorem.

Theorem 1.71–B. *Let X_0 be a proper subspace of the linear space X, and suppose $x_1 \in X - X_0$. Then, there exists an element $x_1' \in X^f$ such that $\langle x, x_1' \rangle = 0$ if $x \in X_0$ and that $\langle x_1, x_1' \rangle = 1$.*

PROOF. We observe that this is the same as Theorem 1.61–D with the hypothesis of finite dimensionality omitted. To prove the theorem, let M be the subspace of X spanned by the set consisting of X_0 and x_1. Elements of M are uniquely representable in the form $x + \alpha x_1$, where $x \in X_0$ and α is a scalar. (See the argument at the beginning of the proof of Theorem 1.71–A.) Let us define $f(x + \alpha x_1) = \alpha$. Then f is a linear functional on M, and $f(x) = 0$ if $x \in X_0$, whereas $f(x_1) = 1$. By Theorem 1.71–A (with Y the linear space of scalars) there exists an element of X^f which is an extension of f. If we denote this element by x_1', we see that it has the required properties.

Theorem 1.71–C. *Let X be a linear space. If $x \in X$ and if $\langle x, x' \rangle = 0$ for every $x' \in X^f$, then $x = 0$. In other words, J^{-1} exists, where J is defined as in* § 1.6 [*see* (1.6–10)].

PROOF. Applying Theorem 1.71–B with $X_0 = (0)$ we see that if $x_1 \neq 0$, there exists $x_1' \in X^f$ such that $\langle x_1, x_1' \rangle \neq 0$. This statement is merely the contrapositive form of Theorem 1.71–C, so the proof is complete. Note that this proposition is the same as Theorem 1.61–B with the hypothesis of finite dimensionality omitted.

Next we consider an extension theorem for linear functionals which are subjected to a special kind of condition.

Definition. Let X be a real linear space, and let p be a real-valued function defined on X, with the properties

$$\begin{aligned} &a.\quad p(x + y) \leqslant p(x) + p(y),\\ &b.\quad p(\alpha x) = \alpha p(x) \quad \text{if } \alpha \geqslant 0. \end{aligned}$$

We call p a *sublinear* functional on X.

As an example of a sublinear functional we cite the example of the space $X = R_n$ with p defined by $p(x) = |\xi_1| + \cdots + |\xi_n|$, where $x = (\xi_1, \cdots, \xi_n)$. Another example is that of $X = C[a, b]$, with $p(x) = |x(a)|$ or $p(x) = \max |x(t)|$ for $a \leqslant t \leqslant b$.

Theorem 1.71–D. *Let X be a real linear space, and let M be a proper subspace of X. Let p be a sublinear functional defined on X and let f be a linear functional defined on M such that $f(x) \leqslant p(x)$ for each $x \in M$. Then, there exists a linear functional F defined on X such that F is an extension of f and $F(x) \leqslant p(x)$ for each $x \in X$.*

PROOF. We have arranged the notation so that the proof can be given in a somewhat abbreviated form by referring to the proof of Theorem

1.71–A. We start with x_0 and M_0 as in that proof. The first problem is to show that there exists a real number ξ_0 such that the linear functional F_0 defined on M_0 by

$$F_0(x + \alpha x_0) = f(x) + \alpha\xi_0 \tag{1.71–1}$$

satisfies the inequality

$$F_0(x + \alpha x_0) \leqslant p(x + \alpha x_0) \tag{1.71–2}$$

for every real α and every $x \in M$.

By taking $\alpha = 1$ in (1.71–1) and (1.71–2) we see that one necessary condition which must be satisfied by ξ_0 is the inequality

$$\xi_0 \leqslant p(x + x_0) - f(x), \tag{1.71–3}$$

which must hold for every $x \in M$. Another similar necessary condition is obtained by taking $\alpha = -1$ and replacing x by $-x$ in (1.71–1) and (1.71–2). In this way we see that the inequality

$$-p(-x - x_0) - f(x) \leqslant \xi_0 \tag{1.71–4}$$

must hold for every $x \in M$.

Now, conversely, if (1.71–3) and (1.71–4) are valid inequalities for every $x \in M$, it follows that (1.71–2) is always satisfied. To see this, suppose $\alpha > 0$, and replace x by $\alpha^{-1}x$ in (1.71–3). Then

$$\xi_0 \leqslant p\left(\frac{x}{\alpha} + x_0\right) - f\left(\frac{x}{\alpha}\right) = \frac{1}{\alpha}p(x + \alpha x_0) - \frac{1}{\alpha}f(x);$$

and so, rearranging,

$$f(x) + \alpha\xi_0 \leqslant p(x + \alpha x_0)$$

for $x \in M$ and $\alpha > 0$. If $\alpha < 0$, replace x by $\alpha^{-1}x$ in (1.71–4). Then

$$\xi_0 \geqslant -p\left(-\frac{x}{\alpha} - x_0\right) - f\left(\frac{x}{\alpha}\right) = \frac{1}{\alpha}p(x + \alpha x_0) - \frac{1}{\alpha}f(x),$$

$$-p(x + \alpha x_0) + f(x) \leqslant -\alpha\xi_0,$$

and so

$$f(x) + \alpha\xi_0 \leqslant p(x + \alpha x_0)$$

if $x \in M$ and $\alpha < 0$. These two results together show that (1.71–2) is always true (the case $\alpha = 0$ is obvious).

Now, suppose x_1 and x_2 are any elements of M. Then

$$\begin{aligned} f(x_2) - f(x_1) &= f(x_2 - x_1) \leqslant p(x_2 - x_1) \\ &= p[(x_2 + x_0) + (-x_1 - x_0)] \leqslant p(x_2 + x_0) + p(-x_1 - x_0), \end{aligned}$$

whence

$$-p(-x_1 - x_0) - f(x_1) \leqslant p(x_2 + x_0) - f(x_2). \tag{1.71–5}$$

Let

$$c = \sup_{x \in M} \{-p(-x - x_0) - f(x)\},$$

$$C = \inf_{x \in M} \{p(x + x_0) - f(x)\}.$$

We see by (1.71–5) that $c \leqslant C$, both c and C necessarily being finite. There certainly exists a real number ξ_0 such that $c \leqslant \xi_0 \leqslant C$, and it will satisfy (1.71–3) and (1.71–4) by the way in which c and C are defined. We then define F_0 by (1.71–1), and (1.71–2) is satisfied. This completes the first phase in the proof of Theorem 1.71–D.

The proof is completed by use of Zorn's lemma in essentially the same way as in the proof of Theorem 1.71–A. We consider the class P of all linear functionals g with $\mathscr{D}(g) \subset X$ such that M is a proper subspace of $\mathscr{D}(g)$, g is an extension of f, and $g(x) \leqslant p(x)$ when $x \in \mathscr{D}(g)$. The partial ordering of P is defined as before, and the rest of the proof is essentially as before.

1.72 Hamel Bases

The German mathematician Hamel conceived the notion of a "basis" for all real numbers as follows: Let H be a set of real numbers with the properties:

1. If $x_1, \cdots, x_n$ is any finite subset of H and if $r_1, \cdots, r_n$ are rational numbers for which $r_1x_1 + \cdots + r_nx_n = 0$, then $r_1 = \cdots = r_n = 0$.
2. Every real number x can be expressed as a finite linear combination of elements of H, with rational coefficients.

In terms of such a basis, Hamel (**1**) then discussed real functions f satisfying the equation $f(x + y) = f(x) + f(y)$ for all real x, y. If H is a basis in the foregoing sense, let a real function f be defined as follows: If $x \in H$, assign the values of $f(x)$ arbitrarily. Any real x has a unique representation $x = r_1x_1 + \cdots + r_nx_n$, where $x_1, \cdots, x_n$ are in H and $r_1, \cdots, r_n$ are rational (n may vary with x, of course). We then define $f(x) = r_1f(x_1) + \cdots + r_nf(x_n)$, the values $f(x_1), \cdots, f(x_n)$ having already been assigned. With this definition f turns out to satisfy the condition $f(x + y) = f(x) + f(y)$ for every x and y. This procedure gives all possible functions f satisfying this condition.

To show the existence of a basis for all real numbers, Hamel used an argument based on the proposition that every set can be well ordered. This proposition is equivalent to the axiom of choice, or to Zorn's lemma.

Hamel's whole procedure can be adapted to the purpose of showing the nature of all possible linear functionals on a linear space X.

Definition. Let X be a linear space with some nonzero elements. A set $H \subset X$ is called a *Hamel basis* of X if:

1. H is a linearly independent subset of X.
2. The linear manifold spanned by H is all of X.

To show the existence of a Hamel basis, let P be the class whose members are the linearly independent subsets of X. Let P be partially ordered by the relation of set inclusion (i.e., $M \prec N$ if $M, N \in P$ and $M \subset N$).

It is easy to see that P satisfies the conditions of Zorn's lemma. For, if $x_1 \neq 0$, the set consisting of x_1 alone is in P, and, if Q is a completely ordered subset of P, the subset of X obtained by taking the union of all the subsets of X comprised in Q is in P and is an upper bound for Q. Hence P must contain a maximal element, say H. Then the linear manifold spanned by H must be X. For, otherwise, some element x of X is not in the linear manifold spanned by H, and the set consisting of x and the elements of H is linearly independent and contains H as a proper subset, contrary to the maximal character of H. Thus H is a Hamel basis of X.

If $x' \in X^f$, we have $x'(x) = \alpha_1 x'(x_1) + \cdots + \alpha_n x'(x_n)$ whenever $x = \alpha_1 x_1 + \cdots + \alpha_n x_n$, the α's being any scalars. Conversely, suppose H is a Hamel basis of X. Define x' by assigning the value of $x'(x)$ arbitrarily when $x \in H$, and define $x'(x) = \alpha_1 x'(x_1) + \cdots + \alpha_n x'(x_n)$ if x is any element of X, uniquely represented in the form $x = \alpha_1 x_1 + \cdots + \alpha_n x_n$ where $x_1, \cdots, x_n$ are in H and $\alpha_1, \cdots, \alpha_n$ are scalars. Then $x' \in X^f$, as is easily seen.

By using the concept of a Hamel basis, we can show that no infinite-dimensional space is algebraically reflexive. First we observe that, if S is a linearly independent subset of X, there is a Hamel basis H of X such that S is a subset of H. This is shown by use of Zorn's lemma, taking P to be the class of all linearly independent subsets of X which contain S.

Theorem 1.72–A. *If X is infinite dimensional, it is not algebraically reflexive. Consequently (see Theorem 1.61–B), a linear space is algebraically reflexive if and only if it is finite dimensional.*

PROOF. Let $H = \{x_i : i \in I\}$ be a Hamel basis of X. Then I is an infinite index set, and $x_i \neq x_j$ if $i \neq j$. Define $x_i' \in X^f$ by $x_i'(x_i) = 1$, $x_i'(x_j) = 0$ if $i \neq j$. Then the set $\{x_i' : i \in I\}$ is a linearly independent set in X^f. For, if we suppose that $\sum_{\nu=1}^{n} \alpha_\nu x'_{i(\nu)} = 0$, we have $0 = \sum_{\nu=1}^{n} \alpha_\nu x'_{i(\nu)}(x_{i(\mu)}) = \alpha_\mu$ when $\mu = 1, \cdots, n$. Now let H' be a Hamel basis of X^f which contains the set $\{x_i' : i \in I\}$.

Let $\{\beta_i : i \in I\}$ be a set of numbers such that $\beta_i \neq 0$ for infinitely many indices i. Define $x'' \in X^{ff}$ by setting $x''(x_i') = \beta_i$ and $x''(x') = 0$ if $x' \in H'$ but x' is not one of the elements x_i'. Consider an element $x_0'' \in X^{ff}$ of the form $x_0'' = Jx$, $x \in X$. We have $x_0''(x_i') = x_i'(x) = \alpha_i$, where α_i is the coefficient of x_i in the representation of x in terms of the Hamel base H. Since $\alpha_i = 0$ for all but a finite number of indices i, it follows that the set of α_i's cannot be the same as the set of β_i's, and therefore $x'' \neq x_0''$. This shows that X^{ff} contains elements not in the range of J, so that X is not algebraically reflexive.

PROBLEMS

1. If H_1 and H_2 are any two Hamel bases for X, they have the same cardinal number. For each $x \in H_1$ let $H_2(x)$ be the finite set of those elements of H_2 which are needed to represent x by using the basis H_2. Show that, if $y \in H_2$, then $y \in H_2(x)$ for some $x \in H_1$, and hence that

$$H_2 = \bigcup_{x \in H_1} H_2(x).$$

The cardinality argument from here on is like that used in another context later in this book; see the last part of the proof of Theorems 3.2–0, following (3.2–11).

2. The cardinal number of a Hamel basis of X is called the dimension of X. In the case of a space X of infinite dimension, the cardinal of X itself can be shown to be the product $C \cdot \dim X$, where C is the cardinal of the set of all real numbers and $\dim X$ is the dimension of X. Since this product is just the larger of the two factors, the cardinal of X is C if $\dim X < C$ and $\dim X$ if $C \leqslant \dim X$. See Jacobson, **1**, Chapter 9, and Löwig, **1**, p. 20.

3. The dimensions of X and X^f are related as follows, provided that $\dim X$ is infinite:

$$\dim X^f = C^{\dim X}.$$

See the reference to Jacobson in the foregoing problem. It follows that X cannot be algebraically reflexive if $\dim X$ is infinite, for in that case $\dim X < \dim X^f$. We can also write $\dim X^f = 2^{\dim X}$ in this case. For, we have $C = 2^{\aleph_0}$, and $\aleph_0 \aleph = \aleph$ for any infinite cardinal $\aleph$, whence $C^{\aleph} = 2^{\aleph}$.

1.8 The Transpose of a Linear Operator

Suppose X and Y are linear spaces, and let A be a linear operator on X into Y. Let X^f and Y^f be the algebraic conjugates of X and Y respectively. We use the notations introduced in § 1.6. To each $y' \in Y^f$ let us make correspond the element $x' \in X^f$ defined by

$$\langle x, x' \rangle = \langle Ax, y' \rangle, \tag{1.8–1}$$

where x varies over all of X, and let us denote the function so defined by A^T: $A^Ty' = x'$. It is easy to check that A^T is a linear operator on Y^f into X^f. We leave the verifications to the reader. Equation (1.8–1) may now be written in the form

$$(1.8\text{–}2) \qquad \langle x, A^Ty'\rangle = \langle Ax, y'\rangle, \qquad x \in X,\ y' \in Y^f.$$

This formula defines A^T.

Definition. The operator A^T defined by (1.8–2) is called the *transpose* of A.

To see the motivation for this terminology, let us consider the case in which X and Y are finite dimensional. Let X be n-dimensional, with basis $x_1, \cdots, x_n$, and let Y be m-dimensional, with basis $y_1, \cdots, y_m$. Let the corresponding dual bases in X^f and Y^f be $x_1', \cdots, x_n'$ and $y_1', \cdots, y_m'$ respectively (see § 1.61). We know from § 1.4 that the operator A determines, and is determined by, a matrix

$$(1.8\text{–}1) \qquad \left\| \begin{matrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{m1} & \alpha_{m2} & \cdots & \alpha_{mn} \end{matrix} \right\|$$

The basic formulas are [see (1.4–3)]

$$(1.8\text{–}2) \qquad Ax_j = \sum_{i=1}^{m} \alpha_{ij} y_i \qquad j = 1, \cdots, n.$$

Since A^T is a linear operator on Y^f into X^f, it follows that A^T is determined by a certain matrix

$$(1.8\text{–}3) \qquad \left\| \begin{matrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \beta_{21} & \beta_{22} & & \beta_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \beta_{n1} & \beta_{n2} & \cdots & \beta_{nm} \end{matrix} \right\|$$

and the set of equations

$$(1.8\text{–}4) \qquad A^Ty_i' = \sum_{k=1}^{n} \beta_{ki} x_k', \qquad i = 1, \cdots, m.$$

If we use the relations between a basis and the dual basis, we shall be able to find the relation which the matrix (1.8–3) bears to the matrix

(1.8–1). We know that $\langle x_j, x_k' \rangle = \delta_{jk}$ [see (1.61–3)] and likewise that $\langle y_k, y_i' \rangle = \delta_{ki}$. Consequently by (1.8–2),

$$\langle Ax_j, y_i' \rangle = \left\langle \sum_{k=1}^{m} \alpha_{kj} y_k, y_i' \right\rangle = \sum_{k=1}^{m} \alpha_{kj}\, \delta_{ki} = \alpha_{ij}.$$

But also, by (1.8–4),

$$\langle Ax_j, y_i' \rangle = \langle x_j, A^T y_i' \rangle = \left\langle x_j, \sum_{k=1}^{n} \beta_{ki} x_k' \right\rangle = \sum_{k=1}^{n} \beta_{ki}\, \delta_{jk} = \beta_{ji}.$$

Therefore

(1.8–5) $$\alpha_{ij} = \beta_{ji}.$$

In other words, the matrix (1.8–3) has for its k^{th} row the k^{th} column of the matrix (1.8–1). In the customary terminology of matrix algebra, the matrix (1.8–3) is the *transpose* of the matrix (1.8–1). It is on this account that the operator A^T is called the transpose of A.

The main reason for introducing the notion of the transpose of a linear operator is that the discussion of the existence of A^{-1} and the description of the range $\mathscr{R}(A)$ are facilitated by considering the transpose operator A^T. We shall see this in § 1.91.

1.9 Annihilators

The material of this section will be used in the following section, § 1.91, where we discuss certain relations between a linear operator A and its transpose A^T. In the present section we do not consider linear operators. Instead, we are concerned with certain linear manifolds in X and X^f, where X is a linear space and X^f is its algebraic conjugate.

Definition. Suppose S is a nonempty subset of X. Then S^0 denotes the set of all $x' \in X^f$ such that $\langle x, x' \rangle = 0$ if $x \in S$. This set S^0 is called the *annihilator* of S in X^f.

Several obvious facts may be noted:

S^0 is a linear manifold, whether S is or not.

(1.9–1) $$(0)^0 = X^f, \qquad X^0 = (0);$$

(1.9–2) $$S_1 \subset S_2 \text{ implies } S_2{}^0 \subset S_1{}^0.$$

It will be of particular interest to us to consider the annihilator of the range of a linear operator.

If we consider a nonempty subset S of X^f, it will have an annihilator in X^{ff}, namely, the set of all $x'' \in X^{ff}$ such tha[illegible] x'' $= 0$ if $x' \in S$. However, it is also useful to consider another no[illegible] follows:

Definition. If S is any nonempty subset of X^f, we use the symbol ^0S to denote the set of all $x \in X$ such that $\langle x, x' \rangle = 0$ if $x' \in S$. This set ^0S is called the annihilator of S in X.

We observe the facts:

^0S is a linear manifold, whether S is or not.

$$^0(0) = X, \qquad ^0(X^f) = (0). \tag{1.9–3}$$

$$S_1 \subset S_2 \text{ implies } {}^0S_2 \subset {}^0S_1. \tag{1.9–4}$$

The relation $^0(X^f) = (0)$ in (1.9–3) follows from Theorem 1.71–C.

If $S \subset X^f$, there is a very simple relation between the annihilator of S in X and the annihilator of S in X^{ff}. If we think of X as a subspace of X^{ff}, via the canonical imbedding (see § 1.6), then ^0S is just that part of S^0 which lies in X. If we use the canonical imbedding operator J, the exact relationship between S^0 and ^0S is expressed in the formula

$$J(^0S) = J(X) \cap S^0. \tag{1.9–5}$$

The verification of this formula is left to the reader. It is an immediate consequence of the definitions.

The first important theorem about annihilators is as follows; it concerns the annihilator of an annihilator:

Theorem 1.9–A. *If S is a nonempty subset of X, then $^0(S^0)$ is the linear manifold spanned by S. In particular, if M is a subspace in X, we have*

$$^0(M^0) = M. \tag{1.9–6}$$

PROOF. Let M be the linear manifold spanned by S. If $x = \alpha_1 x_1 + \cdots + \alpha_n x_n$, where $x_1, \cdots, x_n \in S$, and if $x' \in S^0$, we have

$$\langle x, x' \rangle = \alpha_1 \langle x_1, x' \rangle + \cdots + \alpha_n \langle x_n, x' \rangle = 0,$$

so that $x \in {}^0(S^0)$. Thus $M \subset {}^0(S^0)$. If we suppose $x_0 \in {}^0(S^0) - M$, we know by Theorem 1.71–B that there exists an element $x_0' \in X^f$ such that $\langle x, x_0' \rangle = 0$ if $x \in M$, but $\langle x_0, x_0' \rangle \neq 0$. The first of these two conditions on x_0' implies that $x_0' \in S^0$, because $S \subset M$. But then $\langle x_0, x_0' \rangle = 0$, because $x_0 \in {}^0(S^0)$. We have now reached a contradiction. Therefore $^0(S^0) \subset M$, and the proof is complete.

Corollary 1.9–B. *If* M_1 *and* M_2 *are subspaces of* X *such that* $M_1{}^0 = M_2{}^0$, *it follows that* $M_1 = M_2$.

PROOF. By (1.9–6) we have $M_1 = {}^0(M_1{}^0) = {}^0(M_2{}^0) = M_2$.

In case of finite-dimensional spaces we have:

Theorem 1.9–C. *Let* X *be a linear space of dimension* n, *and let* M *be a subspace of* X, *of dimension* m. *Then the annihilator* M^0 *has dimension* $n - m$.

PROOF. We assume that $n > 1$ and that M is different from (0) and from X, so that $0 < m < n$. The proof is very simple otherwise, and we leave the verification to the reader in these other cases. Let $x_1, \cdots, x_n$ be a basis for X such that $x_1, \cdots, x_m$ is a basis for M. Let $x_1', \cdots, x_n'$ be the dual basis of X^f. Suppose $x \in M$. Then $x = \xi_1 x_1 + \cdots + \xi_n x_n$, where $\xi_i = 0$ if $m < i \leqslant n$. Hence, if $m < j \leqslant n$, $\langle x, x_j' \rangle = \xi_1 \langle x_1, x_j' \rangle + \cdots + \xi_m \langle x_m, x_j' \rangle = 0$, because of the relations (1.61–3). This shows that $x_j' \in M^0$. On the other hand, if $x' \in M^0$, we have $x' = \xi_1' x_1' + \cdots + \xi_n' x_n'$ and, if $1 \leqslant i \leqslant m$, $0 = \langle x_i, x' \rangle = \xi_1' \langle x_i, x_1' \rangle + \cdots + \xi_n' \langle x_i, x_n' \rangle = \xi_i'$, again by (1.61–3). Thus x' is in the linear manifold spanned by $x_{m+1}', \cdots, x_n'$. These two results together show that $x_{m+1}', \cdots, x_n'$ constitute a basis for M^0, which therefore has dimension $n - m$.

Next we consider the situation analogous to that of Theorem 1.9–A when we start with a subset of X^f.

Theorem 1.9–D. *Suppose* S *is a nonempty subset of* X^f, *and let* M *be the linear manifold spanned by* S. *Then*

$$M \subset ({}^0S)^0. \tag{1.9–7}$$

The proof is simple; we leave it to the reader.

Unlike the situation in Theorem 1.9–A, it can happen that $M \neq ({}^0M)^0$ when M is a subspace of X^f. An example will be given presently.

Theorem 1.9–E. *Let* M *be a proper subspace of* X^f. *Then* $M = ({}^0M)^0$ *if and only if to each* $x' \in X^f - M$ *corresponds some* $x \in {}^0M$ *such that* $\langle x, x' \rangle \neq 0$.

PROOF. By (1.9–7) we know that $M \subset ({}^0M)^0$. Hence $({}^0M)^0 = M$ if and only if $({}^0M)^0 \subset M$. For a given $x' \in X^f$ the statement "to x' corresponds some $x \in {}^0M$ such that $\langle x, x' \rangle \neq 0$" is equivalent to the statement "x' is not in $({}^0M)^0$." Thus the theorem can be rephrased as follows: $({}^0M)^0 \subset M$ if and only if $x' \in X^f - M$ implies $x' \in X^f - ({}^0M)^0$. In this form the assertion is obviously true, since a set-theoretic inclusion relation i reversed when we pass from the sets to their complements.

Definition. A subspace M of X^f is called *algebraically saturated* if $M = ({}^0M)^0$.

We see by (1.9–3) and (1.9–1) that (0) and X^f are both algebraically saturated subspaces of X^f. Also, it follows from (1.9–6) that, if M is any subspace of X, its annihilator M^0 is an algebraically saturated subspace of X^f.

Theorem 1.9–F. *If M_1 and M_2 are algebraically saturated subspaces of X^f and ${}^0M_1 = {}^0M_2$, it follows that $M_1 = M_2$.*

PROOF. $M_1 = ({}^0M_1)^0 = ({}^0M_2)^0 = M_2$.

Theorem 1.9–G. *If X is finite dimensional, every subspace of X^f is algebraically saturated.*

PROOF. Let M be a subspace of X^f, we can assume $M \neq X^f$, since we already know that X^f is saturated. Suppose $x_0' \in X^f - M$. By Theorem 1.61–D there exists an element $x_0'' \in M^0$ such that $x_0''(x_0') = 1$. But X is algebraically reflexive, and so there is an element $x_0 \in X$ such that $Jx_0 = x_0''$. Then $\langle x_0, x_0' \rangle = \langle x_0', x_0'' \rangle = 1$. The fact that $x_0'' \in M^0$ implies that $\langle x_0, x' \rangle = \langle x', x_0'' \rangle = 0$ if $x' \in M$; therefore $x_0 \in {}^0M$. Consequently M is algebraically saturated, by Theorem 1.9–E.

We now give an example in which $M \neq ({}^0M)^0$.

Example 1. Let X be the space consisting of all convergent sequences $x = \{\xi_n\}$, with addition and scalar multiplication defined in the natural way. Let M be the subspace of X^f consisting of all linear functionals x' expressible in the form

$$(1.9\text{–}8) \qquad \langle x, x' \rangle = \sum_{i=1}^{\infty} \xi_i \xi_i' \qquad \text{where } \sum_{i=1}^{\infty} |\xi_i'| < \infty.$$

Let x_0' be the particular element of X^f defined by

$$(1.9\text{–}9) \qquad \langle x, x_0' \rangle = \lim_{n\to\infty} \xi_n.$$

We shall show that $x_0' \in ({}^0M)^0 - M$. First we show that $({}^0M)^0 = X^f$, so that certainly $x_0' \in ({}^0M)^0$. Suppose that $x \in {}^0M$, i.e., that $\langle x, x' \rangle = 0$ for each $x' \in M$. For any given positive integer n let $\xi_i' = 0$ if $i \neq n$, $\xi_n' = 1$, and let x' be the corresponding element of M, as in (1.9–8). Then $0 = \langle x, x' \rangle = \xi_n$; this is true for $n = 1, 2, \cdots$, and so $x = 0$. This shows that ${}^0M = (0)$; it follows that $({}^0M)^0 = X^f$. Next we show that x_0' is not in M. For, if we suppose $x_0' \in M$, we have a representation

$$(1.9\text{–}10) \qquad \langle x, x_0' \rangle = \sum_{i=1}^{\infty} \xi_i \alpha_i, \qquad \sum_{i=1}^{\infty} |\alpha_i| < \infty.$$

For a given k let x be chosen with $\xi_n = 0$ if $n \neq k$, $\xi_k = 1$. Then, comparing (1.9–9) and (1.9–10), we see that $\alpha_k = 0$. This holds for each k, so (1.9–10) tells us that $\langle x, x_0' \rangle = 0$ for every x. This is false, for (1.9–9) shows that $\langle x, x_0' \rangle = 1$ when x is defined by $\xi_1 = \xi_2 = \cdots = 1$. We have thus shown that $x_0' \in ({}^0M)^0 - M$. Therefore M is not algebraically saturated.

1.91 Ranges and Null Manifolds

In this section we consider an arbitrary linear operator A on X into Y, where X and Y are linear spaces with the same field of scalars. In § 1.3 we discussed in general terms two important questions which arise in connection with a linear operator: (1) For which elements $y \in Y$ does the equation $Ax = y$ have a solution x? (2) Is the solution unique in all cases in which a solution exists? The first question leads us to seek various ways of characterizing the range $\mathscr{R}(A)$. The second question is the same as the question: "Does the inverse A^{-1} exist?" In the present section we investigate these questions with the aid of the transpose operator A^T (see § 1.8) and the annihilator concepts developed in § 1.9.

Definition. By the *null manifold* of A is meant the set of all $x \in X$ such that $Ax = 0$. We denote this set by $\mathscr{N}(A)$.

It is evident that $\mathscr{N}(A)$ *is a linear manifold in* X.

The definition of null manifold can be applied to the transpose A^T, so that $\mathscr{N}(A^T)$ is the set of all $y' \in Y^f$ such that $A^T y' = 0$.

We recall that the inverse A^{-1} exists if and only if $Ax = 0$ implies $x = 0$, that is, if and only if $\mathscr{N}(A) = (0)$.

Our first results deal with the relations between $\mathscr{R}(A)$ and $\mathscr{N}(A^T)$.

Theorem 1.91–A. $\{\mathscr{R}(A)\}^0 = \mathscr{N}(A^T)$.

PROOF. By definition of A^T we have $\langle Ax, y' \rangle = \langle x, A^T y' \rangle$ for $x \in X$, $y' \in Y^f$. If $y' \in \{\mathscr{R}(A)\}^0$ we see that $\langle Ax, y' \rangle = 0$ for each $x \in X$, and so $A^T y' = 0$, or $y' \in \mathscr{N}(A^T)$. The reasoning is reversible, and so the theorem is proved.

Theorem 1.91–B. $\mathscr{R}(A) = {}^0\{\mathscr{N}(A^T)\}$.

PROOF. $\mathscr{R}(A) = {}^0\{\mathscr{R}(A)^0\} = {}^0\{\mathscr{N}(A^T)\}$, by Theorems 1.9–A and 19.1–A.

Theorem 1.91–C. $\mathscr{R}(A) = Y$ *if and only if* $(A^T)^{-1}$ *exists.*

PROOF. $\mathscr{R}(A) = Y$ implies $\mathscr{N}(A^T) = Y^0 = (0)$, by Theorem 1.91–A and (1.9–1). But $\mathscr{N}(A^T) = (0)$ is equivalent to the existence of $(A^T)^{-1}$. Thus, $\mathscr{R}(A) = Y$ implies the existence of $(A^T)^{-1}$. Conversely, $\mathscr{N}(A^T) = (0)$ implies $\mathscr{R}(A) = {}^0(0) = Y$, by Theorem 1.91–B and (1.9–3).

The foregoing theorems give us what we want with respect to $\mathscr{R}(A)$. Next we seek to obtain information about $\mathscr{N}(A)$ by considering $\mathscr{R}(A^T)$.

Theorem 1.91–D. $\mathscr{N}(A) = {}^0\{\mathscr{R}(A^T)\}$.

PROOF. The formula $\langle x, A^Ty'\rangle = \langle Ax, y'\rangle$, valid for $x \in X$ and $y' \in Y^f$, shows at once that $\mathscr{N}(A) \subset {}^0\{\mathscr{R}(A^T)\}$. The same formula shows that $x \in {}^0\{\mathscr{R}(A^T)\}$ implies $\langle Ax, y'\rangle = 0$ for every $y' \in Y^f$. But then $x \in \mathscr{N}(A)$ by Theorem 1.71–C. This completes the proof.

Theorem 1.91–E. *We always have* $\mathscr{R}(A^T) \subset \{\mathscr{N}(A)\}^0$. *We have* $\mathscr{R}(A^T) = \{\mathscr{N}(A)\}^0$ *if* $\mathscr{R}(A^T)$ *is algebraically saturated. This is certainly the case if* X *is finite dimensional.*

PROOF. By Theorem 1.91–D we have $\{{}^0\{\mathscr{R}(A^T)\}\}^0 = \{\mathscr{N}(A)\}^0$. Then $\mathscr{R}(A^T) \subset \{\mathscr{N}(A)\}^0$, by Theorem 1.9–D. For the rest of the theorem we use the definition of "algebraically saturated" and Theorem 1.9–G.

Theorem 1.91–F. *If* $\mathscr{R}(A^T) = X^f$, *then* A^{-1} *exists.*

PROOF. This follows by Theorem 1.91–D and (1.9–3).

Theorem 1.91–G. *If* A^{-1} *exists, it follows that* $\{{}^0\mathscr{R}(A^T)\}^0 = X^f$. *If* $\mathscr{R}(A^T)$ *is algebraically saturated (hence certainly if* X *is finite dimensional), we can conclude further that* $\mathscr{R}(A^T) = X^f$.

PROOF. If A^{-1} exists, $\mathscr{N}(A) = (0)$. Then $\{{}^0\{\mathscr{R}(A^T)\}\}^0 = (0)^0 = X^f$ by Theorem 1.91–D and (1.9–1). The rest of the theorem follows at once (see Theorem 1.91–E).

We emphasize that, if X is finite dimensional, A^{-1} exists if and only if $\mathscr{R}(A^T) = X^f$.

PROBLEM

If X^f is algebraically reflexive, so is X. Prove this without resort to questions of finite dimensionality, using the following suggestions: Let J_0 be the canonical mapping of X into X^{ff}, J_1 the canonical mapping of X^f into X^{fff}. By Theorem 1.91–C, X is algebraically reflexive if and only if $(J_0{}^T)^{-1}$ exists. By hypothesis $\mathscr{R}(J_1) = X^{fff}$. Show that $J_1{}^{-1} = J_0{}^T$, whence $(J_0{}^T)^{-1}$ exists $(= J_1)$.

1.92 Conclusions

The theorems of § 1.91 involve the ranges and null manifolds of the operator A and its transpose A^T. If we suppose that the spaces X, Y and the operator A come to us as known things, the theorems will be effectively useful to us in furnishing information about $\mathscr{R}(A)$ and $\mathscr{N}(A)$ only to the extent that we know or can find out the relevant information about $\mathscr{R}(A^T)$ and $\mathscr{N}(A^T)$. For this, we must, at the very least, know some things about the spaces X^f, Y^f and the operator A^T.

In the case in which X and Y are finite dimensional, the spaces X^f and Y^f are also finite dimensional. In this case the operator A is fully represented by a matrix, and the transpose A^T is represented by the transposed matrix. For the finite-dimensional case all the results of § 1.91 are well known in algebra (though perhaps in a different terminology) as part of the theory of systems of linear equations or of the theory of matrices.

For infinite-dimensional spaces the results of § 1.91 are not as useful as could be wished, because the algebraic conjugate of an infinite-dimensional space is not as amenable to study as the algebraic conjugate of a finite-dimensional space. The theory for finite-dimensional spaces is much simplified by the fact that we have simple "concrete representations" of such spaces. If X is an n-dimensional linear space with real scalars, it is isomorphic to the arithmetic space R_n (see the proof of Theorem 1.4–A), and we can regard R_n as a concrete representation of X. Likewise, if A is a linear operator on X into Y (spaces of dimensions n and m respectively), we have a concrete representation of A by a certain $m \times n$ matrix. The purpose of such a representation is to replace one object of study by another object whose properties are more readily apparent because the structure of the object is more familiar or more susceptible to intuitive perception.

In order to obtain results of the type presented in § 1.91, but more useful, it seems essential to abandon the purely algebraic approach to linear problems for infinite-dimensional spaces and to bring in topological considerations. If we impose a specified topological structure on a linear space X, it is natural to consider those linear functionals on X which are continuous with respect to the topology. These form a subspace of the algebraic conjugate space X^f; this subspace may be called the *topological conjugate* of X. In a good many interesting special cases the topological conjugate of a given topological linear space X can be studied more effectively than the algebraic conjugate of X. This may be the case, for one thing, because the topological conjugate is significantly smaller than the algebraic conjugate (i.e., has fewer elements). It may also turn out that there is a useful concrete representation of the topological conjugate space. Thus, for example, if X is the space of real-valued continuous functions defined on $0 \leqslant t \leqslant 1$, with a certain natural topology for X, the topological conjugate of X is isomorphic to a certain space whose elements form a subclass of the class of all real-valued functions of bounded variation defined on $0 \leqslant t \leqslant 1$; see Theorem 4.32–C.

This necessarily somewhat vague and general discussion is intended to suggest in part why the abstract approach to linear problems, as far as infinite-dimensional spaces are concerned, is more fruitful if suitable topological as well as algebraic methods are employed. There are still other reasons why it is useful to take account of a topology for a given linear space. A given linear operator may have important properties which find their appropriate expression in terms of the topological or metric notions of "nearness," "small variations," and other related ideas.

The greatest successes of the abstract approach to linear problems have been achieved in the study of Hilbert space. A Hilbert space is a very special type of linear space with a topology. It has many of the properties of Euclidean space and is, therefore, amenable to the use of geometric reasoning guided by intuition. But perhaps the decisive reason for the importance of Hilbert space, and for the greater success which has been achieved with Hilbert space than with other spaces, is the fact that the topological conjugate of a Hilbert space is isomorphic with the given Hilbert space.

SUGGESTED READING

Banach, **1**, especially Chapter 2.
Bourbaki, **1**, especially p. 37 in connection with Zorn's lemma.
Bourbaki, **3**.
Halmos, **1**.
Hamburger and Grimshaw, **1**.
Jacobson, **1**, especially Chapters 1, 2, 9.
Kelley, **1**, especially pp. 31–36 in connection with Zorn's lemma.
Zaanen, **1**, especially Chapter 8.

2

TOPOLOGIES

2.0 Scope of the Chapter

This chapter is intended primarily for reference purposes. It contains the definitions and theorems relating to general topological matters which are needed at various places throughout the book. Since this is not a book on general topology, the selection of topics and the treatment of them conform to our primary interest in linear spaces.

Although the reader is not presumed to have made an extensive study of abstract topology as a prerequisite for studying this book, some acquaintance with the ideas of abstract topology is needed. The student with some experience in topology should proceed directly to Chapter 3 after a preliminary scanning of Chapter 2. Thereafter he can refer back to Chapter 2 as needed. The student who has never studied abstract general topology will probably find it best to read parts of Chapters 2 and 3 in close conjunction, paying especial attention to the examples in Chapter 3. It is not necessary to master all of these chapters before going on to later chapters.

2.1 Topological Spaces

The very general concept of a topological space has evolved by abstraction from various special situations; the original model for a topological space is the real axis, i.e., Euclidean 1-dimensional space. The basic notion in the general definition of a topological space is that of axiomatizing the properties of open sets. A topological space is a set of elements in which certain subsets are designated as "open sets." The class of these open sets must satisfy certain axioms.

Definition. A set X with a family $\mathscr{T}$ of subsets is called a *topological space* if $\mathscr{T}$ satisfies the following conditions:

(a) The empty set $\emptyset$ and the whole space X belong to $\mathscr{T}$.
(b) The union of any number of members of $\mathscr{T}$ (even uncountably many) is again a member of $\mathscr{T}$.
(c) The intersection of any finite number of members of $\mathscr{T}$ is again a member of $\mathscr{T}$.

The family $\mathscr{T}$ is called a *topology* for X, and the members of $\mathscr{T}$ are called the *open sets* of X in this topology.

Example 1. If X is any nonempty set and $\mathscr{T}$ consists of $\emptyset$ and X, $\mathscr{T}$ is evidently a topology for X.

Example 2. If X is any nonempty set and $\mathscr{T}$ consists of *all* subsets of X ($\emptyset$ and X are of course included), $\mathscr{T}$ is evidently a topology for X. It is called the *discrete* topology for X.

Example 3. Let X be the set of all real numbers x. A subset T of X will be called open if to each x_0 in T corresponds some $h > 0$ such that $x \in T$ if $|x - x_0| < h$. Evidently X is open, and $\emptyset$ satisfies vacuously the condition of being open. The axioms (a), (b), (c) for open sets are satisfied. The topology thus defined for X is the one regularly used in analysis.

It is clear from Examples 1 and 2 that the general definition of a topology for a set X leaves the way open for situations quite different from those encountered in the customary topology of the real axis or of the Euclidean plane.

A set S in the topological space X is called *closed* if its complement S' is open. It will be seen that the empty set $\emptyset$ and the whole space X are both open and closed. Because of the complementation rules, it follows that any intersection of closed sets is closed and that any finite union of closed sets is closed. Also, since $(S')' = S$, a set is open if and only if its complement is closed.

If $S \subset X$, the intersection of all closed subsets of X which contain S is called the *closure* of S and denoted by $\bar{S}$. Clearly $\bar{S}$ is closed and $S \subset \bar{S}$. Moreover, it is easy to see that $S = \bar{S}$ if and only if S is closed.

If $S \subset X$, the union of all the open subsets of S is called the *interior* of S and denoted by int (S). Evidently int (S) is open and int $(S) \subset S$. Moreover, S is open if and only if $S = \text{int}\,(S)$.

The *boundary* of a set S is defined as the intersection of the closures of S and its complement S'.

Any open set which contains a point x is called a *neighborhood* of x. Likewise, if $S \subset X$, an open set U such that $S \subset U$ is called a neighborhood of S.

A point x is called an *accumulation point* of a set S if every neighborhood of x contains a point of S distinct from x. It is not hard to see that a set S is closed if and only if it contains all its accumulation points.

When we set out to construct a topology for a given set we often achieve this end by selecting certain sets which turn out to be open and which are used in determining all the other open sets. For example, in our definition of the topology of the real axis, in Example 3, we made use of the *open intervals* with a given point x_0 as center. This idea finds its general expression in the concept of a *base*.

A family $\mathscr{B}$ of open sets is called a *base at the point* x if $x \in B$ whenever $B \in \mathscr{B}$ and if to each neighborhood U of x corresponds some $B \in \mathscr{B}$ such that $B \subset U$. A family of open sets is called a *base for the space* X (or simply a base) if for each $x \in X$ there is a subfamily which is a base at x. A family $\mathscr{B}$ is thus a base for X if and only if to each $x \in X$ and each neighborhood U of x there is a set $B \in \mathscr{B}$ such that $x \in B$ and $B \subset U$. The subfamily of all elements of $\mathscr{B}$ which are neighborhoods of x is then a base at x. Clearly the same family $\mathscr{B}$ cannot be a base for two different topologies on X.

Example 4. Consider the topology for the axis of real numbers defined in Example 3. The set of all open intervals of lengths $2/n$ ($n = 1, 2, \cdots$) is a base for this topological space, and the open intervals $\{x: |x - x_0| < 1/n\}$ ($n = 1, 2, \cdots$) form a base at x_0.

If $\mathscr{B}$ is a base for a topological space X, the following assertions are evidently true:

(a) To each $x \in X$ corresponds some $B \in \mathscr{B}$ such that $x \in B$.
(b) If $B_1, B_2 \in \mathscr{B}$ and $x \in B_1 \cap B_2$, there exists some $B_3 \in \mathscr{B}$ such that $x \in B_3 \subset (B_1 \cap B_2)$.

Now let us suppose that X is a set, initially without any specified topology, and suppose that there is a family $\mathscr{B}$ of subsets of X satisfying the two conditions a, b just given. Let us then *define* a set $S \subset X$ to be open if for each $x \in S$ there is some $B \in \mathscr{B}$ with $x \in B \subset S$. The class

$\mathscr{T}$ of open sets is a topology for X. The verification is easy. The empty set $\emptyset$ is open, since it satisfies the condition vacuously. The set X is open, by (a). Condition (b) assures that finite intersections of open sets are open, and unions of open sets are seen to be open by the definition. The members of $\mathscr{B}$ are themselves open, according to the definition, and $\mathscr{B}$ is a base for the topological space X whose topology $\mathscr{T}$ was defined by means of $\mathscr{B}$. This illustrates an important method of introducing a topology on a set.

Example 5. In R_n (Example 1, § 1.2) let $\mathscr{B}$ be the family of sets B of the following sort: Choose an arbitrary point $x_0 = (\xi_1{}^0, \cdots, \xi_n{}^0)$ and an arbitrary positive number r; let B be the set of points $x = (\xi_1, \cdots, \xi_n)$ such that $(\xi_1 - \xi_1{}^0)^2 + \cdots + (\xi_n - \xi_n{}^0)^2 < r^2$. This family $\mathscr{B}$ has the properties (a) and (b) mentioned in the preceding discussion; the topology for R_n determined by $\mathscr{B}$ as a base in the standard topology for R_n.

Sometimes we consider several topologies for the same set X. If two topologies $\mathscr{T}_1$ and $\mathscr{T}_2$ for X are such that every member of $\mathscr{T}_1$ is also a member of $\mathscr{T}_2$ (i.e., $\mathscr{T}_1 \subset \mathscr{T}_2$), we say that $\mathscr{T}_1$ is *weaker* than $\mathscr{T}_2$ and that $\mathscr{T}_2$ is *stronger* than $\mathscr{T}_1$. The stronger topology has more open sets. Other terminologies in this situation are: $\mathscr{T}_1$ is *coarser* than $\mathscr{T}_2$ and $\mathscr{T}_2$ is *finer* than $\mathscr{T}_1$; also, $\mathscr{T}_2$ is *larger* than $\mathscr{T}_1$.

If $\mathscr{F}$ is any nonempty family of topologies for X, the intersection of all these topologies is a topology for X, and it is weaker than all the topologies in $\mathscr{F}$. Suppose that $\mathscr{A}$ is any nonempty class of subsets of X. There certainly is a topology for X which includes the members of $\mathscr{A}$, namely the discrete topology for X. Hence there is a well-defined topology for X which is the intersection of all those topologies for X which include the members of $\mathscr{A}$. We call it *the topology generated by* $\mathscr{A}$. The open sets in this topology are: $\emptyset$, X, and all unions of finite intersections of members of $\mathscr{A}$. The family $\mathscr{A}$ is called a *subbase* for the topology which it generates.

In our definition of a topological space there are no *separation* axioms. It can happen that, given two distinct points x_1, x_2, every neighborhood of x_1 contains x_2. This implies that x_1 is an accumulation point of the set consisting of x_2 alone and, thus, that the latter set is not closed. There are various separation axioms that may be imposed on a topological space. One such axiom provides that, if x_1 and x_2 are distinct points, there exists an open set S_1 such that S_1 contains x_1 but not x_2. This is equivalent to the requirement that every set consisting of a single point be closed.

Definition. A topological space is called a T_1-space if each set consisting of a single point is closed.

A more stringent separation axiom requires that, if $x_1 \neq x_2$, there shall exist open sets S_1, S_2 such that $x_1 \in S_1$, $x_2 \in S_2$, and $S_1 \cap S_2 = \emptyset$. Spaces with this property are called T_2-spaces, or Hausdorff spaces. They are discussed in § 2.3.

PROBLEMS

1. If $S \subset X$ and $x \in X$, $x \in \bar{S}$ if and only if every neighborhood of x intersects S.

2. A point x is in the boundary of a set S if and only if every neighborhood of x intersects both S and S'.

3. ~~A set S~~ is open if and only if to each point x in S corresponds some neighborhood U of x such that $U \subset S$.

4. The closure of a set consists of the set together with all its accumulation points.

5. If $\mathscr{B}$ is a base for X, every open set in X is the union of certain (perhaps uncountably many) members of the base.

6. If $\mathscr{T}$ is the discrete topology for X (see Example 2), every subset of X is both open and closed. The discrete topology is stronger than every other topology.

7. The weakest possible topology for X is that one in which consists solely of $\emptyset$ and X.

8. In the xy-plane let $\mathscr{A}$ be the family of sets A of the following sort: Each A is determined by two distinct parallel lines and consists of all points of the xy-plane lying strictly between these two lines. What is the nature of the topology generated by $\mathscr{A}$? In particular, how is this topology related to the topology defined in Example 5 if we regard the xy-plane as the space R_2?

2.11 Relative Topologies

Suppose X is a topological space with topology $\mathscr{T}$, and let X_0 be a subset of X. Then we can define a topology $\mathscr{T}_0$ for X_0 by taking for $\mathscr{T}_0$ the family of sets of the form $T \cap X_0$, where $T \subset \mathscr{T}$. The topology $\mathscr{T}_0$ is called the *relative topology for* X_0 *induced by* (*or inherited from*) *the topology* $\mathscr{T}$.

When it is clear from the context in a certain situation that we are talking about the topology of X and the relative topology $\mathscr{T}_0$ for X_0, a subset of X_0 which is open (or closed) in the topology $\mathscr{T}_0$ is called *relatively open* (or *relatively closed*). A set $S \subset X_0$ is relatively closed if and only if it is of the form $A \cap X_0$, where A is closed. For, S is relatively closed if and only if its complement relative to X_0 is relatively open, so that S itself must be of the form $X_0 - (T \cap X_0)$, where T is open. But

$$X_0 - (T \cap X_0) = X_0 \cap (T \cap X_0)' = X_0 \cap (T' \cup X_0') \\ = (X_0 \cap T') \cup (X_0 \cap X_0') = X_0 \cap T'.$$

Since T' is closed, this shows that S has the form asserted.

It is useful to observe that, if X_0 is a closed subset of X and if $S \subset X_0$ is relatively closed, then S is closed in the topology of X. We can see the truth of this as follows: If $S \subset X_0$, we have $X_0' \subset S'$, and hence $X_0' \cap S' = X_0'$ so that

$$S' = (S' \cap X_0') \cup (S' \cap X_0) = X_0' \cup (S' \cap X_0).$$

But, if S is relatively closed, we have $X_0 - S = X_0 \cap S' = T \cap X_0$, where T is open. Thus

$$S' = X_0' \cup (T \cap X_0) = X_0' \cup T,$$

whence S' is open (for X_0' is open) and S is closed.

2.12 Continuous Functions

Let X and Y be topological spaces, and let f be a function with domain X and range Y. If $S \subset X$, we denote by $f(S)$ the set of all $f(x) \in Y$ corresponding to $x \in S$. If $T \subset Y$, we denote by $f^{-1}(T)$ the set of all $x \in X$ such that $f(x) \in T$. There is no implication in this notation that an inverse function exists.

Definition. The function f is said to be continuous at the point $x_0 \in X$ if to each neighborhood V of $f(x_0)$ in Y there corresponds a neighborhood U of x_0 in X such that $f(U) \subset V$. An equivalent rewording of this definition is convenient for some purposes: f is continuous at x_0 if, for each neighborhood V of $f(x_0)$, the set $f^{-1}(V)$ contains a neighborhood of x_0. We further say that f is continuous on X if it is continuous at each point of X.

Theorem 2.12–A. *Suppose f is a function with domain X and range Y, where X and Y are topological spaces. Then f is continuous on X if and only if $f^{-1}(V)$ is an open set in X whenever V is an open set in Y.*

PROOF. Suppose $f^{-1}(V)$ is open whenever V is open. Let x_0 be any point in X and V any neighborhood of $f(x_0)$. Then $f^{-1}(V)$ is open and contains x_0, so that $f^{-1}(V)$ is a neighborhood of x_0. Thus f is continuous at x_0, by the second form of the definition. For the argument in the other direction, suppose f is continuous on X, and let V be an open set in Y. Then, if $x \in f^{-1}(V)$, V is a neighborhood of $f(x)$, and hence $f^{-1}(V)$ must include a neighborhood of x. This implies that $f^{-1}(V)$ is open.

If f has domain X and range Y and $T \subset Y$, it is easy to see that

$$f^{-1}(T') = [f^{-1}(T)]'. \tag{2.12-1}$$

From this formula and Theorem 2.12–A we see that *f is continuous on X if and only if $f^{-1}(S)$ is a closed set in X whenever S is a closed set in Y.*

Sometimes we have the situation that X and Y are topological spaces and f is a function with domain $\mathscr{D}(f) \subset X$ and range $\mathscr{R}(f) \subset Y$. In this case continuity of f at a point of $\mathscr{D}(f)$ is defined by considering $\mathscr{D}(f)$ and $\mathscr{R}(f)$ as spaces with the relative topologies inherited from X and Y respectively.

Suppose that X and Y are topological spaces and that f is a function with domain X and range Y such that the inverse function f^{-1} exists. If both f and f^{-1} are continuous on their domains, f is called a *homeomorphism* of X onto Y, and X and Y are said to be *homeomorphic*.

2.2 Compact Sets

Let X be a set, S a subset of X, and $\mathscr{F}$ a family of subsets of X such that each point of S belongs to at least one member of $\mathscr{F}$. Then the family $\mathscr{F}$ is said to *cover* S, or to be a *covering* of S. If X is a topological space and if all the sets in $\mathscr{F}$ are open, $\mathscr{F}$ is called an *open covering* of S.

Definition. A subset S of a topological space X is called *compact* if every open covering of S includes a finite subfamily which covers S.

It is readily seen that S is compact as a subset of X if and only if S is compact when we consider S itself to be the whole space and use the relative topology.

The empty set is compact. So is any finite set.

Theorem 2.2–A. *A set $S \subset X$ is compact if and only if in every family of relatively closed subsets of S, the intersection of all of which is $\emptyset$, there is a finite subfamily with this same property.*

PROOF. A family of relatively open sets covers S if and only if their complements (relative to S) form a family of relatively closed sets with intersection $\emptyset$. The theorem follows from this remark.

Theorem 2.2–B. *A closed subset of a compact space is compact.*

PROOF. Let X be a compact space and S a closed subset of X. Let $\mathscr{F}$ be a family of relatively closed subsets of S whose intersection is $\emptyset$. If $F \in \mathscr{F}$, F is closed as a subset of X (this follows by an observation made in § 2.11). Hence, by Theorem 2.2–A (applied to the compact set X), $\mathscr{F}$ includes a finite subfamily whose intersection is $\emptyset$. Hence, again by Theorem 2.2–A, S is compact.

Theorem 2.2–C. *Let f be a continuous function whose domain is compact. Then its range is also compact.*

PROOF. We can assume that the domain of f is all of X and the range is all of Y. If $\mathscr{V}$ is a family of open sets V and if $\mathscr{V}$ covers Y, the family $\mathscr{U}$ of sets $U = f^{-1}(V)$ certainly covers X. Moreover, each U is open, by Theorem 2.12–A. Since X is compact, $\mathscr{U}$ includes a finite subfamily $U_1, \cdots, U_n$ which covers X. The corresponding finite subfamily $V_1, \cdots, V_n$ clearly covers Y, and so Y is compact.

We remind the reader that a point set on the real axis (or in n-dimensional Euclidean space) is compact if and only if it is closed and bounded. It follows from Theorem 2.2–C that, if f is real valued and continuous on a compact topological space, the values of f are bounded. Moreover, the set of values is closed, so that the function actually attains its least upper bound and its greatest lower bound. In other words, the function has a maximum value at some point of the space, and likewise a minimum value.

Sometimes we consider sets S with the property that the closure $\bar{S}$ is compact. A set of this kind is called *conditionally* compact (or, also, *relatively* compact).

2.21 Category. Separability

Suppose X is a topological space. If $S \subset X$ and $\bar{S} = X$, we say that S is *dense* in X. If the interior of $\bar{S}$ is empty, we say that S is *nowhere dense.*

Definition. A set $S \subset X$ is said to be of the *first category* if S is the union of a countable family of nowhere dense sets. If S is not of the first category, it is said to be of the *second category.* If $S \subset X$ and $x \in X$, we say that S is of the first category at x if there is a neighborhood U of x such that $U \cap S$ is of the first category; otherwise $U \cap S$ is of the second category for every neighborhood U of x, and S is said to be of the second category at x.

It can be verified that if S is of the first category it is of the first category at every point of X, and conversely.

Definition. A topological space X is called *perfectly separable* if there exists a countable family of sets forming a base for X (or, briefly, if X has a countable base). A space with a finite base will also be called perfectly separable.

The Euclidean plane, for example, is perfectly separable. As a countable base we may take the family of all interiors of circles with rational radii and rational coordinates for the centers.

If the space X is perfectly separable, it contains an at most countable set S which is dense in X. In fact, if $\mathscr{B}$ is a finite or countable base for X and if S is a set formed by selecting one element from each of the sets in $\mathscr{B}$, then S is at most countable and $\bar{S} = X$. For, if we assume $X - \bar{S}$ to be nonempty, it is an open set containing some point x, and hence (since $\mathscr{B}$ is a base) there exists some $B \in \mathscr{B}$ with $x \in B \subset X - \bar{S}$. Some point of B must belong to S, however, and this is a contradiction, since $B \cap \bar{S} = \emptyset$.

Definition. A topological space X is called *separable* if there exists a finite or countable set S which is dense in X.

The foregoing argument shows that a perfectly separable space is separable. The converse is not true in general, though it is true for certain types of spaces (e.g., metric spaces; see § 2.4).

Some spaces, though not perfectly separable, have the property that *at each point* there is a countable base. The requirement that this be so is called the *first axiom of countability*. The requirement of perfect separability is called the *second axiom of countability*.

2.3 Separation Axioms. Hausdorff Spaces

In § 2.1 we gave the definition of a T_1-space: a topological space in which sets consisting of a single point are closed. We now define several other types of spaces, each of which is characterized by the addition of one or more separation axioms to the axioms for a topological space.

Before stating these definitions we remark that two sets A, B are called *disjoint* if $A \cap B = \emptyset$.

Hausdorff Spaces. A topological space is called a *Hausdorff space* if for each pair of distinct points x_1, x_2 in the space there exists disjoint neighborhoods U_1, U_2 of x_1 and x_2 respectively.

Regular Spaces. A topological space is called *regular* if for each point x and each closed set A not containing x there exists disjoint neighborhoods of x and A respectively. An equivalent formulation: For each point x

and each neighborhood U of x there exists a neighborhood V of x such that $\bar{V} \subset U$.

Normal Spaces. A topological space is called *normal* if for each pair of disjoint closed sets A, B there exist disjoint neighborhoods U, V of A and B respectively.

Every Hausdorff space is a T_1-space, but some T_1-spaces are not Hausdorff spaces. A space can be regular or normal without being a T_1-space. However, as applied to T_1-spaces, the types Hausdorff, regular, normal, form a hierarchy: Every normal space is regular and every regular space is a Hausdorff space. There is a nomenclature according to which spaces are classified as T_1, T_2, T_3, T_4. The definitions for T_2, T_3, T_4 are as follows:

T_2-space: a Hausdorff space.
T_3-space: a regular T_1-space.
T_4-space: a normal T_1-space.

The class of T_3-spaces is a proper subclass of the class of T_2-spaces. Likewise for T_4 and T_3.

Compact Hausdorff spaces of a rather general sort play an important role in certain parts of functional analysis. Nearly all of the topologies which are used in connection with linear spaces are such as to make the spaces regular; they are usually assumed to be T_1-spaces also, so that they are T_3-spaces.

Later on we shall need certain facts about Hausdorff spaces and normal spaces. We present these facts here as theorems.

Theorem 2.3–A. *If C_0 and C_1 are disjoint compact sets in a Hausdorff space X, there exist disjoint neighborhoods W_0 and W_1 of C_0 and C_1 respectively.*

PROOF. We suppose at first that C_0 consists of a single point x_0 (such a set, in fact any finite set, is compact). Then for every point x in C_1 there exist disjoint neighborhoods U and V of x_0 and x respectively. The family of all the V's covers C_1, and hence some finite subfamily $V_1, \cdots, V_n$ covers C_1. Let $U_1, \cdots, U_n$ be the corresponding U's and let $W_0 = U_1 \cap \cdots \cap U_n$, $W_1 = V_1 \cup \cdots \cup V_n$. This gives what is wanted for this special case. In the general case suppose x is any point of C_0. By what has just been proved there exist disjoint neighborhoods U, V of x and C_1 respectively. The family of all the U's covers C_0, and hence some finite subfamily $U_1, \cdots, U_m$ also covers C_0. If $V_1, \cdots, V_m$ are the corresponding V's, we let $W_0 = U_1 \cup \cdots \cup U_m$, $W_1 = V_1 \cap \cdots \cap V_m$ and attain the desired result.

Theorem 2.3–B. *A compact subset of a Hausdorff space is closed.*

PROOF. Suppose C is compact and $x \in \bar{C}$. Then $x \in C$. For, if not, Theorem 2.3–A implies the existence of a neighborhood of x which doesn't intersect C. This in turn implies that $x \notin \bar{C}$ (see § 2.1). From this contradiction we conclude that $\bar{C} \subset C$; hence $\bar{C} = C$, and C is closed.

A Hausdorff space is sometimes defined (as was originally done by Hausdorff) by a separate set of axioms. Let X be a set of elements, and suppose there is given a family $\mathscr{N}$ of ordered pairs (N, x), where N is a subset of X and $x \in N$, satisfying the following axioms:

1. For each $x \in X$ there is at least one $N \subset X$ such that $(N, x) \in \mathscr{N}$.
2. If (N_1, x) and (N_2, x) are in $\mathscr{N}$, there exists an $(N_3, x) \in \mathscr{N}$ with $N_3 \subset N_1 \cap N_2$.
3. If $(N_1, x_1) \in \mathscr{N}$ and $x_2 \in N_1$, there exists an $N_2 \subset X$ such that $(N_2, x_2) \in \mathscr{N}$ and $N_2 \subset N_1$.
4. If $x_1 \neq x_2$, there exist sets N_1, N_2 in X with (N_1, x_1) and (N_2, x_2) belonging to $\mathscr{N}$, and $N_1 \cap N_2 = \emptyset$.

We now define a set S to be open if to each $x \in S$ there corresponds an element $(N, x) \in \mathscr{N}$ with $N \subset S$. Axioms 1 and 2 guarantee that the open sets thus defined actually yield a topology for X. Axiom 3 assures us that all the sets N occurring in the elements of $\mathscr{N}$ are open. They evidently form a base for X. Finally, axiom 4 guarantees us that the space is a Hausdorff space. Conversely, a Hausdorff space satisfies axioms 1–4 if we take for the elements of $\mathscr{N}$ all the pairs (U, x), where U is an open set containing x.

There is a property called *complete regularity* which a topological space may possess and which is useful in the expression of certain ideas.

Definition. A space X is called *completely regular* if for each $x_0 \in X$ and each neighborhood U of x_0 there can be defined on X a real-valued continuous function f, all of whose values lie on the closed interval $[0, 1]$, such that $f(x_0) = 0$ and $f(x) = 1$ if $x \in X - U$.

A completely regular space is certainly regular. For, if x_0, U, and f are as in the definition of complete regularity, the set $\{x : f(x) < \frac{1}{2}\}$ is a neighborhood of x_0 whose closure lies in U.

Every T_4-space is completely regular. This fact is implied by the following more general result:

Theorem 2.3–C. *Let E and F be disjoint closed sets in a T_4-space X. Then, there can be defined on X a real-valued continuous function f, all of whose values lie on the closed interval* $[0, 1]$, *such that* $f(x) = 0$ *if* $x \in E$ *and* $f(x) = 1$ *if* $x \in F$.

This theorem is commonly known as Urysohn's lemma. We omit the proof in order to save space for other things. For proofs see Loomis, **1**, page 6, or Kelley, **1**, page 115.

We also observe the following:

Theorem 2.3–D. *A compact Hausdorff space is normal.*

PROOF. An immediate consequence of Theorems 2.3–A and 2.2–B.

2.31 Locally Compact Spaces

Definition. A topological space is called *locally compact* if every point has a neighborhood whose closure is compact.

Any topological space X can be imbedded in another topological space Y, having just one more point than X, in such a way that Y is compact and the relative topology of X as a subset of Y is just the original topology of X. It is done in this way: Let y_0 be any element distinct from the points of X. Let $\mathscr{W}$ be the class of open sets W in X such that W' is compact. Observe that $X \in \mathscr{W}$. Let Y be the set consisting of the points of X and the point y_0. A set in Y will be called open if either (*a*) it does not contain y_0 and is open as a subset of X, or (*b*) it does contain y_0 and its intersection with X is a member of $\mathscr{W}$. These sets in Y meet the requirements for defining a topology; we leave verification to the reader. It is evident that the relative topology of X coincides with its original topology.

Now Y is compact. For, supposing $\mathscr{V}$ to be a family of open sets which covers Y, there must be some member of $\mathscr{V}$ of the form $W_0 \cup (y_0)$, where $W_0 \in \mathscr{W}$. Then W_0' is compact as a subset of X. It is covered by the sets $V \cap X$, V varying over $\mathscr{V}$. Hence some finite set $V_1 \cap X, \cdots, V_n \cap X$ covers W_0'. Then $V_1, \cdots, V_n$ and $W_0 \cup (y_0)$ cover Y, so that Y is compact.

The space Y constructed in this way is called the *one-point compactification* of X.

If X is a T_1-space, so is Y. It can happen that Y is not a Hausdorff space, even though X is. However, if X is a locally compact Hausdorff space, Y will be a Hausdorff space. To verify that Y fulfills the Hausdorff separation condition in this case, it is enough to consider pairs of points of which one is y_0, for, if both points are in X, the separation is accomplished by using the same sets as if we were dealing entirely with the space X. If $x \in X$, there exists a neighborhood U of x such that $\bar{U}$ is compact. Then $\bar{U}' \cup (y_0)$ is a neighborhood of y_0 which is disjoint from U, and the separation requirement is satisfied.

It is easy to see, conversely, that, if X and Y are both Hausdorff spaces, then X is locally compact.

Theorem 2.31-A. *Let X be a locally compact Hausdorff space, and let C_0, C_1 be disjoint compact sets in X. Then, there exist disjoint neighborhoods W_0 and W_1 of C_0 and C_1 respectively such that the closures of W_0 and W_1 are compact.*

We state this theorem formally for later reference. Its proof differs but slightly from that of Theorem 2.3–A. One need only observe that, if U and V are neighborhoods of a point and V has compact closure, $U \cap V$ is a neighborhood of the point having the same property. One can then make sure that all the neighborhoods obtained in the earlier proof have compact closures.

Theorem 2.31-B. *Let X be a locally compact Hausdorff space. Let C be a compact set in X, and U a neighborhood of C. Then, there can be defined on X a real-valued continuous function f such that $0 \leqslant f(x) \leqslant 1$ for every x, $f(x) = 1$ if $x \in C$, and $f(x) = 0$ if $x \in U'$.*

PROOF. Let Y be the one-point compactification of X, and in Y let $E = U' \cup (y_0)$, $F = C$. Then E and F are disjoint closed sets in Y, and we can apply Theorem 2.3–C, since Y is normal.

2.4 Metric Spaces

If X is any set, we denote by $X \times X$ the set of all ordered pairs (x_1, x_2), where x_1 and x_2 are in X. A function d with domain $X \times X$ and range in the real number field is called a *distance function* on X if it satisfies the conditions:

(a) $d(x_1, x_2) = d(x_2, x_1)$.
(b) $d(x_1, x_3) \leqslant d(x_1, x_2) + d(x_2, x_3)$.
(c) $d(x_1, x_2) = 0$ if and only if $x_1 = x_2$.
(d) $d(x_1, x_2) \geqslant 0$.

A set X with a distance function is called a *metric space*. We note, incidentally, that (d) is a consequence of (a), (b), and (c) [put $x_1 = x_3$ in (b)]. The number $d(x_1, x_2)$ is called the distance between x_1 and x_2. Property (c) is called the triangular law of distances.

With each point x_0 in a metric space X and with each positive number r we associate the set of all $x \in X$ such that $d(x, x_0) < r$. We call this set the *sphere* with center x_0 and radius r. In some spaces a sphere may contain no points other than the center. The set $\mathscr{N}$ of all pairs (N, x), where N is any sphere with center at x and x is any point of X, satisfies the four Hausdorff axioms (see § 2.3) and thus defines a topology for the metric space, the set of all spheres being a base. The space is, moreover, a Hausdorff space. Whenever we speak of a metric space, it is understood that its topology is defined in this way. Verification that the four Hausdorff axioms are satisfied is left to the reader.

A metric space has a countable base at each point (the set of spheres with positive rational radii and centers at the point). The space is perfectly separable if and only if there exists a finite or countable set which is dense in the space. We already know that such a set exists if the space is perfectly separable (see § 2.21). On the other hand, if we suppose that an at most countable set exists which is dense in the space, the set of all spheres with positive rational radii and centers at the points of this dense set provides a countable base for the space, as is readily verified. Thus, for a metric space, the notions of perfect separability and separability are equivalent.

A sequence $\{x_n\}$ of points in a metric space X is called *convergent* if there is a point $x \in X$ such that $d(x_n, x) \to 0$ as $n \to \infty$; this x is unique if it exists. We then write $x_n \to x$ and call x the *limit* of the sequence.

Theorem 2.4–A. *In a metric space a set S is closed if and only if the situation: $x_n \in S$, $x_n \to x$ implies that $x \in S$.*

PROOF. If $x_n \to x$, where $\{x_n\}$ is a sequence in S, every neighborhood of x intersects S, and so $x \in \bar{S}$. If S is closed, $\bar{S} = S$, and so $x \in S$. On the other hand, suppose S has the property that $x \in S$ whenever $x_n \to x$ and all the x_n are in S. Choose any $x \in \bar{S}$. Then, for each n the sphere of center x and radius $1/n$ intersects S at least in one point, say x_n. Then $d(x_n, x) < 1/n$, and so $x_n \to x$, whence $x \in S$. Thus $\bar{S} \subset S$, and S is closed.

Theorem 2.4–B. *In a metric space a point x is an accumulation point of a set S if and only if there exists a convergent sequence $\{x_n\}$ of distinct points of S with x as limit.*

The proof of this theorem is left to the reader.

Sometimes a set X can be made into a metric space in different ways by different definitions of a distance function. For example, if X is the xy-plane, any one of the expressions

$$[(x-a)^2 + (y-b)^2]^{\frac{1}{2}},$$
$$|x-a| + |y-b|,$$
$$\max\{|x-a|, |y-b|\},$$

can be taken as the definition of the distance between (x, y) and (a, b). It turns out that the topology for X determined by any one of these distance functions is the same as that determined by either of the others.

If d_1 and d_2 are two distance functions on a set X and if $\mathscr{T}_1$ and $\mathscr{T}_2$ are the resulting topologies for X, it is easily proved that $\mathscr{T}_1 = \mathscr{T}_2$ if and only if, for all $\{x_n\}$, $d_1(x_n, x) \to 0$ is equivalent to $d_2(x_n, x) \to 0$.

If X and Y are metric spaces and f is a function with domain $\mathscr{D}(f)$ in X and range $\mathscr{R}(f)$ in Y, the topological definition of continuity (in § 2.12) is equivalent to the familiar $\epsilon\delta$-formulation of continuity: f is continuous at $x_0 \in \mathscr{D}(f)$ if to each $\epsilon > 0$ corresponds a $\delta > 0$ such that $d(f(x), f(x_0)) < \epsilon$ whenever $x \in \mathscr{D}(f)$ and $d(x, x_0) < \delta$. Here we have used the same symbol d for the distance functions in X and Y. Just as in classical analysis, we have the proposition: *f is continuous at x_0 if and only if $f(x_n) \to f(x_0)$ whenever $x_n \to x_0$, $\{x_n\}$ being a sequence of points in $\mathscr{D}(f)$.*

The concept of uniform continuity can be defined for a function with domain and range in metric spaces: f is uniformly continuous on its domain if to each $\epsilon > 0$ there corresponds a $\delta > 0$ such that $d(f(x_1), f(x_2)) < \epsilon$ whenever $x_1, x_2 \in \mathscr{D}(f)$ and $d(x_1, x_2) < \delta$.

A subset S in a metric space X can be considered as a metric space by itself, using the same distance function as that defined on X, but confining attention to distances between pairs of points in S. The resulting topology of the metric space S is the same as the relative topology of S induced by the topology of X.

A set in a metric space is called *bounded* if it is contained in some sphere.

Theorem 2.4–C. *In a metric space a compact set is closed and bounded.*

PROOF. We already know that a compact set is closed, since a metric space is a Hausdorff space (see Theorem 2.3–B). Hence we have only to prove that a compact set is bounded. We may assume the set is nonempty. If the compact set is S and if $x \in S$, let V_n be the (open) sphere of center x and radius n. The family of all the V_n's covers S (in fact, covers the space). Since S is compact, a finite number of the V_n's suffice to cover S, and hence $S \subset V_N$ for some index N. Thus S is bounded.

Theorem 2.4–D. *If S is a compact set in a topological space, every infinite subset of S has an accumulation point in S.*

PROOF. In this theorem the space need not be metric, nor even a Hausdorff space. Suppose T is an infinite subset of S but that T has no accumulation point in S. Then to each $x \in S$ there corresponds some neighborhood U of x which intersects T at most in the one point x (if x happens to be in T). The set of all these neighborhoods covers S, and so a finite number of them, $U_1, \cdots, U_n$, suffice to cover S. Since

$T \subset U_1 \cup \cdots \cup U_n$ and at most one point of T is in each U_k, it follows that T is finite. This contradiction completes the proof.

Theorem 2.4–E. *If S is a subset of a metric space X with the property that every infinite subset of S has an accumulation point in $\overset{S}{X}$, there exists a countable family $\mathcal{N}$ of open sets in X such that, if U is any open set in X and $x \in S \cap U$, there is some set $N \in \mathcal{N}$ for which $x \in N \subset U$.*

PROOF. We assert that for each positive integer n there is some finite set of points $x_1, \cdots, x_{N(n)}$ in S such that the set of spheres of radii $1/n$ with centers at $x_1, \cdots, x_{N(n)}$ covers S. For, if this were not so, there would be some n and a sequence of points $x_1, x_2, \cdots$ in S such that $d(x_i, x_j) \geqslant 1/n$ if $i \neq j$. The infinite set of all these x's could not have any accumulation point, contrary to the hypothesis on S. Now let $\mathcal{N}$ be the countable family of all these finite sets of spheres. It is easy to see that this family $\mathcal{N}$ has the property asserted for it.

Theorem 2.4–F. *A compact metric space is separable.*

PROOF. We see by Theorems 2.4–D and 2.4–E that, if the metric space X is compact, it has a countable base.

Theorem 2.4–G. *If S is a set in a metric space with the property that every infinite subset of S has an accumulation point in S, then S is compact.*

PROOF. Suppose $\mathcal{U}$ is an open covering of S. Now, Theorem 2.4–E applies to S; hence let $\mathcal{N}$ be a countable family of open sets with the property specified in Theorem 2.4–E. Let $\mathcal{N}^*$ be the subfamily of $\mathcal{N}$ defined as follows: $N \in \mathcal{N}^*$ if and only if $N \in \mathcal{N}$ and there is some $U \in \mathcal{U}$ with $N \subset U$. By the property of $\mathcal{N}$ and the fact that $\mathcal{U}$ covers S, we see that $\mathcal{N}^*$ covers S. Now let $\mathcal{U}^*$ be a subfamily of $\mathcal{U}$ obtained by taking from $\mathcal{U}$ just one U such that $N \subset U$, for each $N \in \mathcal{N}^*$. Then $\mathcal{U}^*$ is at most countable, and it covers S. It remains to prove that some finite subfamily of $\mathcal{U}^*$ covers S. Let the sets in $\mathcal{U}^*$ be indexed: $U_1, U_2, \cdots$. Suppose, contrary to what we wish to prove, that for each n the finite union $U_1 \cup \cdots \cup U_n$ fails to cover S, so that there is some point $x_n \in S - (U_1 \cup \cdots \cup U_n)$. The set of x_n's must be infinite, and hence must have an accumulation point, say x_0, in S. Then $x_0 \in U_N$ for some index N, and by Theorem 2.4–B $x_n \in U_N$ for infinitely many values of n, in particular for some $n > N$. This contradicts the fact that $x_n \in S - (U_1 \cup \cdots \cup U_n)$, and so the proof is complete.

Theorems 2.4–D and 2.4–G together show us that, *in a metric space, a set S is compact if and only if every infinite subset of S has an accumulation point in S.* For many purposes the following criterion for compactness is convenient:

Theorem 2.4–H. *In a metric space, a set S is compact if and only if every sequence in S contains a convergent subsequence with limit in S.*

The proof of this theorem is left to the reader.

Let S be a nonempty set in the metric space X, and let x_0 be any point in X. The distance between S and x_0 is denoted by $d(S, x_0)$ and defined as

$$d(S, x_0) = \operatorname*{glb}_{x \in S} d(x, x_0).$$

If S_1 and S_2 are two nonempty sets in X, the distance between them is denoted by $d(S_1, S_2)$ and defined as

$$d(S_1, S_2) = \operatorname{glb} d(x_1, x_2) \qquad (x_1 \in S_1, x_2 \in S_2).$$

Evidently $d(S_1, S_2) = d(S_2, S_1)$. It is not difficult to see that

$$d(S_1, S_2) = \operatorname*{glb}_{x \in S_2} d(S_1, x) = \operatorname*{glb}_{x \in S_1} d(S_2, x). \tag{2.4–1}$$

For any x_0, $d(x, x_0)$ is a continuous function of x. For, it is readily seen by the properties of the distance function that

$$|d(x, x_0) - d(x', x_0)| \leqslant d(x, x'),$$

whence the asserted continuity follows. It is also true that $d(x_1, x_2)$ is a continuous function of both x_1 and x_2. For, it may be seen that

$$|d(x, x') - d(x_0, x_0')| \leqslant d(x, x_0) + d(x', x_0'). \tag{2.4–2}$$

It follows from the continuity of $d(x, x_0)$ as a function of x that, if S is a nonempty compact set, there is a point $x \in S$ such that $d(S, x_0) = d(x, x_0)$ (see Theorem 2.2–C and the remarks following it). If S is a fixed nonempty set, then $d(S, x)$ is a continuous function of x; we leave the proof of this to the reader. It follows that, if S_1 is any nonempty set and S_2 is a nonempty compact set, there is a point $x_2 \in S_2$ such that $d(S_1, S_2) = d(S_1, x_2)$. Hence, if S_1 and S_2 are both compact, there exist points $x_1 \in S_1$, $x_2 \in S_2$ such that $d(S_1, S_2) = d(x_1, x_2)$. A different proof of this last assertion can be made using Theorem 2.4–H, and we give this proof in the next paragraph to illustrate the convenience of the characterization of compact sets in the last mentioned theorem.

If S_1 and S_2 are any nonempty sets, there exist sequences $\{x_n'\}$, $\{x_n''\}$ in S_1 and S_2 respectively such that

$$d(S_1, S_2) = \lim_{n \to \infty} d(x_n', x_n'').$$

If S_1 is compact, $\{x_n'\}$ contains a convergent subsequence with limit

$x_1 \in S_1$. Let this subsequence be denoted by $\{y_n'\}$, and let the corresponding subsequence of $\{x_n''\}$ be denoted by $\{y_n''\}$. Then

$$d(S_1, S_2) = \lim_{n\to\infty} d(y_n', y_n'').$$

If S_2 is also compact, $\{y_n''\}$ contains a convergent subsequence with limit $x_2 \in S_2$. Let this subsequence be denoted by $\{z_n''\}$, and let the corresponding subsequence of $\{y_n'\}$ be denoted by $\{z_n'\}$. Then $z_n' \to x_1$, $z_n'' \to x_2$, and

$$d(S_1, S_2) = \lim_{n\to\infty} d(z_n', z_n''),$$

whence $d(S_1, S_2) = d(x_1, x_2)$.

If $d(S_1, S_2) > 0$, it follows that $S_1 \cap S_2 = \emptyset$. The converse is not true, however, even if S_1 and S_2 are closed. For example, in the xy-plane, the sets $S_1 = \{(x, y): y \geqslant e^x\}$ and $S_2 = \{(x, y): y \leqslant 0\}$ are closed and disjoint, and yet $d(S_1, S_2) = 0$.

Suppose S is any nonempty set in a metric space. Then $d(S, x) = 0$ if and only if $x \in \bar{S}$, as is easily seen. This result can be generalized as follows:

Theorem 2.4–I. *Suppose that S_1, S_2 are nonempty sets and that S_2 is compact. Then $d(S_1, S_2) = 0$ if and only if $\bar{S}_1 \cap S_2 \neq \emptyset$.*

PROOF. As we saw previously, $d(S_1, S_2) = d(S_1, x_2)$, where x_2 is some point in S_2. Thus $d(S_1, S_2) = 0$ implies $d(S_1, x_2) = 0$, which in turn implies $x_2 \in \bar{S}_1$, so that $\bar{S}_1 \cap S_2 \neq \emptyset$. Conversely, if $x_2 \in \bar{S}_1 \cap S_2$, it follows that $d(S_1, x_2) = 0$, and this implies $d(S_1, S_2) = 0$, by (2.4–1).

Theorem 2.4–J. *A metric space is normal.*

PROOF. Suppose S_1 and S_2 are nonempty closed sets in a metric space. For each point $x \in S_1$ construct the sphere with center x and radius $^1/_2 d(S_2, x)$. Let U_1 be the union of all such spheres as x varies over S_1. Likewise construct a set U_2, related to S_2 as U_1 is related to S_1. Then U_1 and U_2 are open, and clearly $S_k \subset U_k$ $(k = 1, 2)$. It remains only to prove that $U_1 \cap U_2 = \emptyset$. Suppose that $x \in U_1 \cap U_2$. Then, there exist points x_1 and x_2 in S_1 and S_2 respectively such that $d(x, x_1) < {}^1/_2(S_2, x_1)$, $d(x, x_2) < {}^1/_2 d(S_1, x_2)$. Consequently, by the triangular law,

$$d(x_1, x_2) < \tfrac{1}{2}d(S_2, x_1) + \tfrac{1}{2}d(S_1, x_2). \tag{2.4–3}$$

Now

$$d(S_2, x_1) \leqslant d(x_2, x_1) \qquad \text{and} \qquad d(S_1, x_2) \leqslant d(x_1, x_2). \tag{2.4–4}$$

Consequently, we see from (2.4–3) that $d(S_2, x_1) \leqslant d(S_1, x_2)$ and $d(S_1, x_2) \leqslant d(S_2, x_1)$, whence $d(S_2, x_1) = d(S_1, x_2)$. But then (2.4–3) becomes $d(x_1, x_2) < d(S_2, x_1)$, which contradicts (2.4–4). We conclude that $U_1 \cap U_2 = \emptyset$. Therefore the space is normal.

2.41 Completeness

If X is a metric space and $\{x_n\}$ is a convergent sequence in X, with limit x, then $d(x_n, x_m) \leqslant d(x_n, x) + d(x, x_m)$, whence it follows that $d(x_n, x_m) \to 0$ as $m \to \infty$ and $n \to \infty$. It can happen, however, that a sequence $\{x_n\}$ is such that $d(x_n, x_m) \to 0$ as $m \to \infty$ and $n \to \infty$, and yet the sequence has no limit. A sequence $\{x_n\}$ is called a *Cauchy sequence* if $d(x_n, x_m) \to 0$ as m and $n \to \infty$ (i.e., if to each $\epsilon > 0$ corresponds an N such that $d(x_n, x_m) < \epsilon$ when $N \leqslant m$ and $N \leqslant n$).

Definition. If a metric space X has the property that every Cauchy sequence in X has a limit in X, then X is said to be *complete.*

The property of completeness is very important in the construction of existence proofs. The fundamental difference between the fields of real and rational numbers, respectively, is that the real field is complete and the rational field is not when these fields are considered as metric spaces with $d(x_1, x_2) = |x_1 - x_2|$ as the distance function.

Definition. Two metric spaces X and Y are said to be *isometric* if there is a function f with domain X and range Y such that the inverse function f^{-1} exists and furthermore such that $d(x_1, x_2) = d(f(x_1), f(x_2))$ for every pair of points $x_1, x_2 \in X$.

It is a fact of fundamental importance that an incomplete metric space can always be regarded as a dense subset in a complete metric space. The precise statement is as follows:

Theorem 2.41–A. *Let X be an incomplete metric space. There exists a complete metric space Y and a subset Y_0 dense in Y such that Y_0 and X are isometric.*

PROOF. Let $\{x_n\}$ and $\{x_n'\}$ be two Cauchy sequences in X. We write $\{x_n\} \sim \{x_n'\}$ and say that the two sequences are *equivalent* if $d(x_n, x_n') \to 0$ as $n \to \infty$. Equivalence, thus defined, is reflexive, symmetric, and transitive, so that the set of all Cauchy sequences in X is divided into equivalence classes. We denote by Y the set of all these equivalence classes. We shall define a distance function on Y.

First we observe that if $\{x_n\}$ and $\{x_n'\}$ are Cauchy sequences in X, $d(x_n, x_n')$ is a Cauchy sequence of real numbers. For, by (2.4–2)

$$|d(x_n, x_n') - d(x_m, x_m')| \leqslant d(x_n, x_m) + d(x_n', x_m'),$$

whence we see the truth of the assertion just made. Next we assert that $\lim_{n\to\infty} d(x_n, x_n')$ (which exists, since the real numbers form a complete space) is unchanged if we replace $\{x_n\}$ by any equivalent Cauchy sequence, and likewise with $\{x_n'\}$. This is because, if $\{x_n\} \sim \{a_n\}$ and $\{x_n'\} \sim \{a_n'\}$,

$$|d(x_n, x_n') - d(a_n, a_n')| \leqslant d(x_n, a_n) + d(x_n', a_n').$$

We leave the rest of this argument to the reader. Now, if $y, y' \in Y$, let $\{x_n\}$ and $\{x_n'\}$ be Cauchy sequences from the equivalence classes y, y' respectively. We define $d(y, y') = \lim_{n\to\infty} d(x_n, x_n')$. By what has been said, this limit exists and is uniquely determined by y, y'. It is not hard to show that the function d thus defined on $Y \times Y$ is a distance function, so that Y is a metric space. We leave these details to the reader.

Let Y_0 be the subset of Y consisting of those equivalence classes which contain a Cauchy sequence $\{x_n\}$ for which $x_1 = x_2 = x_3 = \cdots$. There cannot be two different Cauchy sequences of this kind in the same equivalence class. If $y, y' \in Y_0$ and if corresponding Cauchy sequences are $\{x_n\}$ and $\{x_n'\}$, with $x_n = x$, $x_n' = x'$ for every n, we see that $d(y, y') = d(x, x')$. It is then clear that Y_0 and X are isometric. Next we show that $\bar{Y}_0 = Y$. If $y \in Y$, let $\{x_n\}$ be a Cauchy sequence from the class y, and suppose $\epsilon > 0$. There exists an integer N such that $d(x_n, x_N) < \epsilon/2$ if $N \leqslant n$. Let $y_0 \in Y_0$ be the class containing the repetitive sequence $x_N, x_N, \cdots$. Then $d(y, y_0) = \lim_{n\to\infty} d(x_n, x_N) \leqslant \epsilon/2 < \epsilon$. This shows that $y \in \bar{Y}_0$, and hence that $Y = \bar{Y}_0$.

Finally, we show that Y is complete. Suppose $\{y_n\}$ is a Cauchy sequence in Y. For each n choose $z_n \in Y_0$ so that $d(y_n, z_n) < 1/n$; we can do this since Y_0 is dense in Y. Now

$$\begin{aligned} d(z_n, z_m) &\leqslant d(z_n, y_n) + d(y_n, y_m) + d(y_m, z_m) \\ &< \frac{1}{n} + \frac{1}{m} + d(y_n, y_m), \end{aligned}$$

and so $\{z_n\}$ is a Cauchy sequence. Let x_n be the element in X which is matched with z_n in the isometric correspondence between X and Y_0. Then $\{x_n\}$ is a Cauchy sequence in X. Accordingly, there is a unique equivalence class $y \in Y$ which contains $\{x_n\}$. Now

$$d(y_n, y) \leqslant d(y_n, z_n) + d(z_n, y) < \frac{1}{n} + d(z_n, y),$$

and since $d(z_n, y) = \lim_{m\to\infty} d(x_n, x_m)$, it is clear that $y_n \to y$. This proves that Y is complete and finishes the proof of the theorem.

In practice we identify X with Y_0 and regard X as a dense set in Y. The space Y is called the *completion* of X. We can, if we choose, adjoin the elements of $Y - Y_0$ to X and form a space isometric to Y in which X is genuinely a dense set. When this is done we shall denote the completion of X by $\hat{X}$.

A nonempty set S in a metric space is said to have finite diameter if $d(x, x')$ is bounded as x and x' vary over S. The diameter of S is then defined as the least upper bound of the distances $d(x, x')$. We denote the diameter of S by diam (S). A nonempty set has finite diameter if and only if it is bounded.

Theorem 2.41–B. *Let X be a complete metric space. Suppose $\{S_n\}$ is a sequence of nonempty closed sets in X such that $S_1 \supset S_2 \supset S_3 \supset \cdots$ and such that* $\operatorname{diam}(S_n) \to 0$ *as $n \to \infty$. Then the intersection of all the sets S_n is nonempty.*

PROOF. This theorem is commonly known as Cantor's intersection theorem. To prove the theorem, choose a point $x_n \in S_n$ for each n. If $m < n$, x_m and x_n are both in S_m and so $d(x_n, x_m) \leqslant \operatorname{diam}(S_m)$. Thus $\{x_n\}$ is a Cauchy sequence, and, since X is complete, the sequence has a limit x in X. It is a simple matter to prove that $x \in S_m$ for every m. We leave the completion of the proof to the reader.

Theorem 2.41–C. *Let X be a nonempty complete metric space. Then X is of the second category as a subset of itself. Moreover, X is of the second category at every one of its points.*

PROOF. We shall refer to this theorem, and especially to the first assertion in the theorem, as Baire's category theorem. If $x_0 \in X$ and $r > 0$, let us write $K(x_0, r) = \{x : d(x, x_0) \leqslant r\}$. This set is closed; we call it the closed sphere with center x_0 and radius r. The set $\{x : d(x, x_0) < r\}$, which we have previously called the sphere of center x_0 and radius r, may for emphasis be referred to as an *open* sphere. It is clear that any open sphere with center x_0 contains a closed sphere with the same center, and vice versa.

Now suppose that X is of the first category, so that X is the union of nowhere dense sets $S_1, S_2, S_3, \cdots$. The fact that S_k is nowhere dense can be expressed by the statement that any closed sphere contains another closed sphere which does not intersect S_k. Hence we see by induction that there exists a sequence of closed spheres $K_n = K(x_n, r_n)$ such that $K_n \cap S_n = \emptyset$, $0 < r_n < 1/n$, and $K_{n+1} \subset K_n$. Now $\operatorname{diam}(K_n) \leqslant 2/n$; hence we can apply Cantor's intersection theorem to assert that there exists a point x in the intersection of all the closed spheres K_n. But then x is not in any S_n and hence not in X, which is a contradiction. Therefore X is of the second category.

To prove the second assertion in the theorem, suppose some point $x \in X$ has a neighborhood V which is of the first category, so that $V = \bigcup_{n=1}^{\infty} S_n$, where S_n is nowhere dense. Let K be a nonempty closed sphere contained in V. Then (see Theorem 2.41–E) K in itself is a complete metric space. Now $K = \bigcup_{n=1}^{\infty} (S_n \cap K)$. It is easily verified that $S_n \cap K$ is nowhere dense in K, and hence K is of the first category in itself, contrary to the fact that it is a complete space. This contradiction completes the proof.

The next theorem has applications in the study of linear operators, as well as elsewhere in analysis.

Theorem 2.41–D. *Let X be a metric space and S a set of the second category in X. Let $\mathscr{F}$ be a family $\{f\}$ of real-valued, continuous functions defined on X such that for each $x \in S$ there exists a real number $m(x)$ with the property that $f(x) \leqslant m(x)$ for every $f \in \mathscr{F}$. Then, there exists a real constant N and an open sphere K in X such that $f(x) \leqslant N$ if $x \in K$ and $f \in \mathscr{F}$.*

PROOF. For each positive integer n let $S_n = \{x : f(x) \leqslant n \text{ when } f \in \mathscr{F}\}$. The continuity of f insures that S_n is closed. The condition on S and $\mathscr{F}$ insures that $S \subset \bigcup_{n=1}^{\infty} S_n$. Let $T_n = S \cap S_n$. Then $\bar{T}_n \subset \bar{S}_n = S_n$, and $S = \bigcup_{n=1}^{\infty} T_n$. Since S is of the second category, at least one of the sets T_n, say T_N, is such that $\bar{T}_N$ contains a nonempty open set and hence contains an open sphere, say K. Then $K \subset S_N$. This means that $f(x) \leqslant N$ if $x \in K$ and $f \in \mathscr{F}$. The proof is now complete.

Theorem 2.41–E. *Let X be a complete metric space, and X_0 a closed set in X. Then X_0, considered as a metric space by itself, is complete.*

The proof is left to the reader.

Theorem 2.41–F. *Let S be a compact set in a metric space. Then S, considered as a metric space by itself, is complete.*

The proof is left to the reader. Use Theorem 2.4–H.

Suppose that S is a set in the metric space X, and let $\hat{X}$ be the completion of X. If we regard S by itself as a metric space, the completion of S, denoted by $\hat{S}$, can be identified with a subset of $\hat{X}$, and we shall suppose this to be done. If $\bar{S}$ is the closure of S in X, it is easily seen that $\bar{S} = \hat{S} \cap X$. The closure of S in $\hat{X}$ is $\hat{S}$.

Definition. A set S is called *precompact* if $\hat{S}$ is compact.

It is readily seen that S and $\bar{S}$ have the same closure in $\hat{X}$, and therefore the same completion. If S is conditionally compact (see § 2.2), i.e., if $\bar{S}$ is compact, then $\bar{S}$ is complete, and so coincides with its completion $\hat{S}$. Hence a conditionally compact set is precompact. If the space X is complete, $\bar{S}$ coincides with $\hat{S}$. Hence in this case (but not in general) a precompact set is conditionally compact.

Theorem 2.41–G. *A nonempty set S in a metric space X is precompact if and only if to each $\epsilon > 0$ corresponds some finite set $x_1, \cdots, x_n$ of points of X such that S is contained in the union of the spheres of radius ϵ with centers at $x_1, \cdots, x_n$.*

PROOF. If $x_0 \in X$, the sphere in $\hat{X}$ with center x_0 and radius ϵ intersects X in the sphere in X with center x_0 and radius ϵ. Consider the set $\mathscr{F}$ of all spheres in $\hat{X}$ with centers in S, each sphere of radius $\epsilon > 0$. If $x \in \hat{S}$, the sphere in $\hat{X}$ of radius ϵ and center x will contain some point $y \in S$, and hence the sphere in $\hat{X}$ of center y and radius ϵ will contain x. Thus $\mathscr{F}$ covers $\hat{S}$. Assuming now that $\hat{S}$ is compact, we conclude that a finite number of the members of $\mathscr{F}$ suffice to cover $\hat{S}$. Hence the intersections of these spheres with X cover S. Thus one half of the theorem is proved. Note that the centers of the spheres can be taken in S.

Now let us assume that S has the property stated in the theorem, and let us prove that $\hat{S}$ is compact. First we shall show that every sequence of elements from S contains a Cauchy subsequence. Starting with a given sequence $\{x_n\}$, we know that there exists a finite number of spheres in X of radius 1, with centers in X, whose union contains S; therefore, some one of these spheres must contain a subsequence from $\{x_n\}$. Denote this subsequence by $x_{11}, x_{12}, x_{13}, \cdots$. Next we consider spheres of radius $\frac{1}{2}$, and conclude that some sphere of radius $\frac{1}{2}$ must contain a subsequence of the foregoing subsequence. We proceed by induction, obtaining sequences $x_{k1}, x_{k2}, x_{k3}, \cdots, (k = 1, 2, \cdots,)$ each a subsequence of its predecessor, and the kth sequence contained in a sphere of radius $1/k$. The diagonal sequence $x_{11}, x_{22}, x_{33}, \cdots$ is then a Cauchy sequence and a subsequence of the original sequence $\{x_n\}$. Now consider a sequence $\{y_n\}$ from $\hat{S}$. Since $\hat{S}$ is closed in $\hat{X}$ and $\hat{X}$ is complete, $\hat{S}$ will be proved compact if we show that $\{y_n\}$ contains a Cauchy sequence. Now, there exists a point $x_n \in S$ such that $d(x_n, y_n) < 1/n$. The sequence $\{x_n\}$ contains a Cauchy subsequence $\{x_{n(k)}\}$, and the corresponding subsequence $\{y_{n(k)}\}$ will also have the Cauchy property, as is readily seen by the triangular inequality. Thus the proof is complete.

2.5 Product Spaces

If X and Y are nonempty sets, the Cartesian *product* set $X \times Y$ is by definition the set of all ordered pairs (x, y), where $x \in X$ and $y \in Y$. If X and Y are topological spaces, there is a standard method of assigning a topology to $X \times Y$. Let $\mathscr{A}$ be the family of subsets of $X \times Y$ of the form $U \times V$, where U and V are any nonempty open sets in X and Y respectively. Then, we take as the standard topology of $X \times Y$ the topology generated by $\mathscr{A}$ (see § 2.1). The open sets in $X \times Y$ are thus $\emptyset$ and unions of members of $\mathscr{A}$. This definition extends in an obvious manner to products $X_1 \times \cdots \times X_n$ of a finite number of topological spaces.

It is sometimes necessary to define the Cartesian product of an infinite number of topological spaces. This is done as follows. Let I be an index set and $\{X_i\}$ $(i \in I)$ a family of sets. The product set $X = \prod_{i \in I} X_i$ is by definition the set of all functions x with domain I and range $\bigcup_{i \in I} X_i$ such that the value of x at i [denoted by $x(i)$] is in X_i. This definition is consistent with the one already given if I is a finite set. If each X_i is a topological space, let $\mathscr{A}$ be the family of product sets of the form $\prod_{i \in I} S_i$, where for some *finite* subset of the indices i, S_i is an arbitrary nonempty open set in X_i, and $S_i = X_i$ for the rest of the indices. The topology of X is then taken to be that generated by $\mathscr{A}$, and with this topology X is called the Cartesian product of the X_i.

If $x \in X$, the value $x(i)$ is called the *coordinate* of x in X_i. Let f_i be the function with domain X and range X_i defined by $f_i(x) = x(i)$. If (for a fixed i) T_i is an open set in X_i, $f_i^{-1}(T_i)$ is the product set $\prod_{j \in I} S_j$ in which $S_i = T_i$ and $S_j = X_j$ if $j \neq i$. This set is open in X, and so f_i is continuous. Actually, it can be shown that the topology we have assigned to X is the weakest topology such that all the functions f_i are continuous. If S is a set in X, the set $f_i(S)$ is called the *projection* of S onto X_i.

It is easily seen that if all the spaces X_i are Hausdorff spaces, their Cartesian product is also a Hausdorff space.

Suppose for a moment that I is the finite set of integers $1, \cdots, n$. Let us write $x(i) = x_i$. The value $f(x)$ of a function defined on X is often written $f(x_1, \cdots, x_n)$. If the values of f are in a topological space Y, f is continuous at $a = (a_1, \cdots, a_n)$ if and only if to each neighborhood V

of $f(a_1, \cdots, a_n)$ there correspond neighborhoods $U_1, \cdots, U_n$ of $a_1, \cdots, a_n$ in $X_1, \cdots, X_n$, respectively, such that $f(x_1, \cdots, x_n) \in V$ if $x_1 \in U_1, \cdots, x_n \in U_n$.

If $X_1, \cdots, X_n$ is a finite set of metric spaces, the product set $X = X_1 \times \cdots \times X_n$ can be made into a metric space in a number of ways. For example, if $x = (x_1, \cdots, x_n)$ and $a = (a_1, \cdots, a_n)$ are points of X, we can define

$$d(x, a) = d_1(x_1, a_1) + \cdots + d_n(x_n, a_n),$$

where d_i is the distance function in X_i. Other workable definitions of distance in X are

$$d(x, a) = \max_{1 \leqslant i \leqslant n} d_i(x_i, a_i)$$

and

$$d(x, a) = [\{d_1(x_1, a_1)\}^2 + \cdots + \{d_n(x_n, a_n)\}^2]^{\frac{1}{2}}.$$

All three of these definitions of distance yield the same topology for X, and it is the same as the topology of the Cartesian product.

Later on in the book we shall need the following theorem, due to Tychonoff:

Theorem 2.5–A. *The Cartesian product of compact spaces is compact.*

For reasons of space we omit the proof of Theorem 2.5–A. For proofs we refer to Loomis, **1**, page 11, and Kelley, **1**, page 143.

SUGGESTED READING

Banach, **1**, especially the Introduction.
Bourbaki, **2**.
Fréchet, **1**.
Hall and Spencer, **1**, especially Chapters 1–4.
Kelley, **1**.
Loomis, **1**, especially Chapter 1.
Ljusternik and Sobolew, **1**, especially Chapter 1.

3

TOPOLOGICAL LINEAR SPACES

3.0 Introduction

The concept of a topological linear space is the result of combining in a suitable way the concept of a linear space and the concept of a topological space. The phrase "in a suitable way" means that we set up a certain relation between two concepts. The essence of this relation is that we require the topology of the space to be such that the algebraic operations native to the linear space are continuous as functions of the elements which enter into these operations. We state this requirement more precisely as follows:

Definition. Let X be a linear space which is also a topological space in the sense of § 2.1. Let A be the space of scalars associated with X, with its usual topology (that of the real or complex number system). Then $x_1 + x_2$ is a function defined on the product space $X \times X$ and αx is a function defined on $A \times X$. We say that X is a *topological linear space* if $x_1 + x_2$ is a continuous function on $X \times X$ and αx is a continuous function on $A \times X$. The product spaces are given the usual topologies for Cartesian products (see § 2.5).

Purely from a logical point of view it might be appropriate to begin our consideration of topological linear spaces with a study of general features common to all such spaces. We might then proceed to classify topological linear spaces in some way and to study the various classifications in some order determined by a scheme of progressive specialization. Our procedure will be quite different from this. We shall begin with consideration of *normed linear spaces.* These are spaces in which the topology is introduced, by means of a metric, in a simple and natural way by assuming that each vector in the linear space has a *length* and that the rules governing the lengths of vectors conform to a few simple and natural geometrical principles. Most of the later part of this book is concerned with the use of normed linear spaces in functional analysis. Some knowledge of topological linear spaces of more general type is advantageous, even if one restricts his attention mainly to normed linear spaces. On this account we have given some attention to the general fundamental notions about topological linear spaces in the latter half of this chapter. In particular, the study of convex sets is made in the context of general topological linear spaces. A good deal of this material is of importance for applications in connection with normed linear spaces, but for various reasons it seems better to present it in the general form as part of the theory of topological linear spaces. The reader who is interested mainly in normed linear spaces need not have an extensive knowledge of the more general theory in order to get what he needs out of §§ 3.4–3.7. The general form of the Hahn-Banach theorem is given as Theorem 3.7–B. The special form of it needed in normed linear spaces occurs in Chapter 4 as Theorem 4.3–A. For locally convex topological linear spaces the appropriate form is that of Theorem 3.8–D.

Kolmogoroff's criterion for normability of a topological linear space is given in Theorem 3.41–D. In § 3.81 are given some of the fundamental notions about weak topologies for linear spaces. The main use of these ideas in the book is in connection with normed linear spaces (see especially, §§ 4.41, 4.61, 4.62).

Of especial interest among normed linear spaces are the inner-product spaces, in which the length, or *norm*, of a vector is defined in terms of a certain function of two vectors, called the *inner product.* If the vectors are x, y, the inner product is denoted by (x, y). It is formally analogous to the dot product of ordinary vector analysis; in fact, an n-dimensional real inner-product space is congruent to the Euclidean space R_n, in which the inner product (= dot product) of $x = (\xi_1, \cdots, \xi_n)$ and $y = (\eta_1, \cdots, \eta_n)$ is

$$(x, y) = \xi_1\eta_1 + \cdots + \xi_n\eta_n.$$

The relation of the norm to the inner product is that expressed in the formula

$$\|x\| = (x, x)^{\frac{1}{2}}.$$

For R_n this gives

$$\|x\| = (\xi_1^2 + \cdots + \xi_n^2)^{\frac{1}{2}};$$

this is the usual expression for the length of x in Euclidean geometry.

A complete, infinite-dimensional inner-product space is called a Hilbert space. The name is in honor of David Hilbert, who studied a particular space of this sort in connection with his investigations of the theory of integral equations.

In some of those sections in this chapter where there are a great many theorems, a statement will be found early in the section calling attention to the principal theorems.

3.1 Normed Linear Spaces

A *norm* on a linear space X is a real-valued function, whose value at x we denote by $\|x\|$, with the properties:

(a) $\|x_1 + x_2\| \leqslant \|x_1\| + \|x_2\|$.
(b) $\|\alpha x\| = |\alpha| \, \|x\|$.
(c) $\|x\| \geqslant 0$.
(d) $\|x\| \neq 0$ if $x \neq 0$.

Actually, property (c) is a consequence of properties (a) and (b). For $\|0\| = 0$ by (b) (put $\alpha = 0$), and from (a) and (b) $0 = \|x - x\| \leqslant \|x\| + \|-x\| = 2\|x\|$, whence $0 \leqslant \|x\|$.

A linear space on which a norm is defined becomes a metric space if we define $d(x_1, x_2) = \|x_1 - x_2\|$, as is easily verified. A linear space which is a metric space in this way is called a *normed linear space*, or a *normed vector space*. If we regard x as a vector, its length is $\|x\|$; the length $\|x_1 - x_2\|$ of the vector difference $x_1 - x_2$ is the distance between the end points of the vectors x_1, x_2.

If f is a function with domain $\mathscr{D}(f) \subset X$ and range $\mathscr{R}(f) \subset Y$, where X and Y are normed linear spaces, continuity of f at $x_0 \in \mathscr{D}(f)$ is expressed by the condition: to each $\epsilon > 0$ corresponds some $\delta > 0$ such that $\|f(x) - f(x_0)\| < \epsilon$ if $x \in \mathscr{D}(f)$ and $\|x - x_0\| < \delta$.

The set $\{x : \|x - x_0\| < r\}$, where $r > 0$, is the (open) sphere of radius r with center x_0; the closure of this set is $\{x : \|x - x_0\| \leqslant r\}$, which we call the closed sphere of radius r with center x_0. By the *surface* of this sphere we mean the set $\{x : \|x - x_0\| = r\}$.

A set S in a normed linear space is bounded if and only if it is contained in some sphere; an equivalent condition is that S be contained in some sphere with center at 0, which means that $\|x\|$ is bounded for $x \in S$.

The field of scalars associated with a normed linear space may be either the real field or the complex field, and the space itself is called real or complex according to which of these scalar fields is used. Definitions and theorems in which no particular specification is made concerning the scalar field are applicable to both real and complex spaces.

We make free use of the definitions and terminology concerning linear spaces as given in Chapter 1. A subspace of a normed linear space is itself a normed linear space. However, a subspace of a normed linear space may or may not be closed, and the distinction between closed and nonclosed subspaces is often important. The terms *subspace* and *linear manifold* are used interchangeably.

A linear space can be a metric space without being a normed linear space. Also, a linear space can have a topology and yet not be a metric space.

Let X be a normed linear space and let A denote the scalar field associated with X (either the real or the complex field). We observe that A itself is a normed linear space, with the absolute value as norm. Vector addition in X is a function on the Cartesian product $X \times X$. This function is continuous, as we see from the inequality

$$\|(x_1 + x_2) - (y_1 + y_2)\| \leqslant \|x_1 - y_1\| + \|x_2 - y_2\|.$$

Multiplication of vectors by scalars is a function on $A \times X$. This function is continuous, as we see from the relations

$$\begin{aligned}\|\alpha x - \alpha_0 x_0\| &= \|\alpha(x - x_0) + (\alpha - \alpha_0)x_0\| \\ &\leqslant |\alpha|\, \|x - x_0\| + |\alpha - \alpha_0|\, \|x_0\|.\end{aligned}$$

As a result of the continuity of $x_1 + x_2$ and αx, it is easy to see that, if M is a linear manifold in X, the closure $\bar{M}$ is also a linear manifold. If S is any set in X, the closure in X of the linear manifold generated by S (as defined in § 1.1) is called the closed linear manifold generated by S. The word "generated" in this definition is sometimes replaced by "determined" or "spanned."

It is also to be observed that $\|x\|$ is a continuous function of x. This follows from the inequality

$$|\, \|x_1\| - \|x_2\| \,| \leqslant \|x_1 - x_2\|,$$

which follows easily from property (a) of the norm.

Definition. Two normed linear spaces X and Y are said to be *isometrically isomorphic* or, more briefly, congruent, if there is a one-to-one correspondence between the elements of X and Y which makes the two spaces isomorphic in the sense defined in § 1.4 and isometric in the sense defined in § 2.41.

In order that X and Y be congruent it is necessary and sufficient that there exist a linear operator T with domain X and range Y, such that T^{-1} exists and $\|Tx\| = \|x\|$ for every $x \in X$. Here we use the same symbol for the norm in X as for the norm in Y. We shall meet examples of congruent spaces when we deal with concrete representations of various linear spaces whose elements are linear operators.

Definition. Two normed linear spaces X and Y are said to be *topologically isomorphic* if there is a linear operator T (with inverse T^{-1}) which establishes the isomorphism of X and Y and which furthermore has the property that T and T^{-1} are continuous on their respective domains. In other words, X and Y are topologically isomorphic provided there is a homeomorphism T of X onto Y which is also a linear operator. For this reason X and Y may be called *linearly homeomorphic.*

As we see later, there exist examples of pairs of spaces which are topologically isomorphic but not congruent.

It is convenient at this point to consider a few facts about continuous linear operators in normed linear spaces. Linear operators, without regard to continuity, were defined in § 1.3.

Theorem 3.1–A. *Let X and Y be normed linear spaces and T a linear operator on X into Y. Then T is continuous either at every point of X or at no point of X. It is continuous on X if and only if there is a constant M such that $\|Tx\| \leqslant M\|x\|$ for every x in X.*

PROOF. Let x_0 and x_1 be any points of X, and suppose T is continuous at x_0. Then to each $\epsilon > 0$ corresponds $\delta > 0$ such that $\|Tx - Tx_0\| < \epsilon$ if $\|x - x_0\| < \delta$. Now suppose $\|x - x_1\| < \delta$. Then $\|(x + x_0 - x_1) - x_0\| < \delta$, and so $\|T(x + x_0 - x_1) - Tx_0\| < \epsilon$. But $T(x + x_0 - x_1) - Tx_0 = Tx - Tx_1$, and so $\|Tx - Tx_1\| < \epsilon$. This shows that, if T is continuous at x_0, it is also continuous at x_1; thus, the first assertion of the theorem is proved.

If $\|Tx\| \leqslant M\,\|x\|$ for all x, it is clear that T is continuous at 0 [recall that $T(0) = 0$]. Conversely, if T is continuous at 0, $\|Tx\|$ can be kept as small as we please by keeping x sufficiently small. Thus, in particular, there is some $\delta > 0$ such that $\|Tx\| < 1$ if $\|x\| < \delta$. Now, if $x \neq 0$, let $x_0 = (\delta x)/(2\|x\|)$, so that $\|x_0\| = \delta/2 < \delta$. Then $\|Tx_0\| < 1$. But $Tx_0 = [\delta/(2\|x\|)]Tx$, and so

$$\|Tx\| = \frac{2\|x\|}{\delta}\,\|Tx_0\| < \frac{2}{\delta}\|x\|.$$

Thus, if we take $M = 2/\delta$, we have $\|Tx\| \leqslant M\|x\|$ if $x \neq 0$, and this inequality evidently is true as well when $x = 0$. The proof is thus complete.

If $\|Tx\| \leqslant M\|x\|$ for every x, it is plain that $\|Tx\| \leqslant M$ when $\|x\| \leqslant 1$. Vice versa, if $\|Tx\| \leqslant M$ for every x such that $\|x\| \leqslant 1$ (or even for every x such that $\|x\| = 1$), then $\|Tx\| \leqslant M\|x\|$ for every x. This is proved by an argument like that used in the last part of the foregoing proof. The smallest admissible value of M in the inequality $\|Tx\| \leqslant M\|x\|$ is called the *bound* of T. It is also called the *norm* of T and denoted by $\|T\|$. The reason for this terminology and notation will appear later. The following formulas for the norm of T are useful:

$$\|T\| = \sup_{\|x\| \leqslant 1} \|Tx\|, \tag{3.1–1}$$

$$\|T\| = \sup_{\|x\| = 1} \|Tx\|, \tag{3.1–2}$$

$$\|T\| = \sup_{x \neq 0} \frac{\|Tx\|}{\|x\|}. \tag{3.1–3}$$

For (3.1–2) and (3.1–3) we assume, of course, that X contains some x such that $x \neq 0$.

When T is a continuous linear operator on X into Y, it is clear from Theorem 3.1–A that a set S lying in the sphere $\|x\| \leqslant r$ in X is carried into the set $T(S)$ lying in the sphere $\|y\| \leqslant \|T\|r$ in Y. Hence $T(S)$ is bounded if S is bounded. Conversely, if T is linear and has the property that $T(S)$ is a bounded set whenever S is a bounded set, it follows that T is continuous. We leave verification of this assertion to the reader.

Theorem 3.1–B. *Suppose T is a linear operator on X to Y, where X and Y are normed linear spaces. Then, the inverse T^{-1} exists and is continuous on* seu *its domain of definition if and only if there exists a constant $m > 0$ such that*

$$m\|x\| \leqslant \|Tx\| \tag{3.1–4}$$

for every $x \in X$.

PROOF. If (3.1–4) holds and $Tx = 0$, it follows that $x = 0$. Then T^{-1} exists, by Theorem 1.3–B. Now $y = Tx$ is equivalent to $x = T^{-1}y$. Hence (3.1–4) is equivalent to $m\|T^{-1}y\| \leqslant \|y\|$, or $\|T^{-1}y\| \leqslant \frac{1}{m}\|y\|$, for all y in the range of T, which is the domain of T^{-1}. This implies that T^{-1} is continuous, by Theorem 3.1–A. We leave the converse proof to the reader.

Theorem 3.1–C. *If X and Y are normed linear spaces, they are topologically isomorphic if and only if there exists a linear operator T with domain X and range Y, and positive constants m, M such that*

$$m\|x\| \leqslant \|Tx\| \leqslant M\|x\| \tag{3.1–5}$$

for every x in X.

PROOF. This theorem is a direct corollary of Theorems 3.1–A and 3.1–B.

Theorem 3.1–D. *Let X be a linear space, and suppose two norms $\|x\|_1$ and $\|x\|_2$ are defined on X. These norms define the same topology on X if and only if there exist positive constants m, M such that*

$$m\|x_1\| \leqslant \|x\|_2 \leqslant M\|x\|_1 \tag{3.1–6}$$

for every x in X.

PROOF. Let X_i be the normed linear space which X becomes with the norm $\|x\|_i$ $(i = 1, 2)$. Let $Tx = x$, and consider T as an operator with domain X_1 and range X_2. Condition (3.1–6) is precisely the condition that T and T^{-1} both be continuous. They are both continuous if and only if the open sets in X_1 are the same as the open sets in X_2 (Theorem 2.12–A). The conclusion now follows.

3.11 Examples of Normed Linear Spaces

In this section we define and establish standard notations for a number of spaces which will be referred to at various places throughout the book. In defining the norms in these various spaces we frequently leave it for the reader to verify that the norm, as defined, actually has the properties required of a norm.

In many of our examples, the elements of the spaces are functions defined on some set T. The values of the functions may be either real or complex; we get a real or complex space according to whether the values of the functions are real or complex. It is always to be understood that, if x_1, x_2 and x are functions defined on T, the functions $x_1 + x_2$ and αx are defined by

$$(x_1 + x_2)(t) = x_1(t) + x_2(t), \qquad (\alpha x)(t) = \alpha x(t), \qquad t \in T.$$

Example 1. The spaces R_n and C_n (defined in Examples 1 and 2, § 1.2) can be made into normed linear spaces in a variety of ways. Let us consider C_n. If $x = (\xi_1, \cdots, \xi_n) \in C_n$ and $p \geqslant 1$, let us define

$$\|x\| = (|\xi_1|^p + \cdots + |\xi_n|^p)^{1/p}. \tag{3.11–1}$$

The triangularity property $\|x_1 + x_2\| \leqslant \|x_1\| + \|x_2\|$ is the same as Minkowski's inequality for finite sums, namely

$$\left(\sum_{i=1}^{n} |\xi_i + \eta_i|^p\right)^{1/p} \leqslant \left(\sum_{i=1}^{n} |\xi_i|^p\right)^{1/p} + \left(\sum_{i=1}^{n} |\eta_i|^p\right)^{1/p}$$

For reference to Minkowski's inequality see the Introduction. When C_n is considered as a normed linear space with the norm (3.11–1), we denote the space by $l^p(n)$. We can make the real space R_n into a normed space in the same way. We shall use $l^p(n)$ for both the real and the complex space with norm (3.11–1); whether the real or complex space is under discussion at a given time will either be clear from the context, or we shall make a specific statement if necessary.

We can also define a norm in R_n and C_n by the formula

$$\|x\| = \max\,[|\xi_1|, \cdots, |\xi_n|].$$

The notation for the space with this norm is $l^\infty(n)$. This notation is natural, since

$$\max_{1\leqslant k\leqslant n} |\xi_k| = \lim_{p\to\infty} (|\xi_1|^p + \cdots + |\xi_n|^p)^{1/p}. \tag{3.11–2}$$

Example 2. We also have spaces analogous to $l^p(n)$ and $l^\infty(n)$ for *infinite* sequences. The space l^p, where $p \geqslant 1$, is defined to consist of all sequences $x = \{\xi_n\}$ such that $\sum_{n=1}^{\infty} |\xi_n|^p < \infty$. The norm in l^p is defined by

$$\|x\| = \left(\sum_{n=1}^{\infty} |\xi_n|^p\right)^{1/p}. \tag{3.11–3}$$

Minkowski's inequality for infinite sums is used in showing that the requisite properties of a norm are fulfilled.

The space l^∞ is defined to consist of the *bounded* sequences $x = \{\xi_n\}$, with the norm

$$\|x\| = \sup_n |\xi_n|.$$

For any $x \in l^\infty$ we have

$$\sup_n |\xi_n| = \lim_{n\to\infty} \left[\lim_{p\to\infty} (|\xi_1|^p + \cdots + |\xi_n|^p)^{1/p}\right]. \tag{3.11–4}$$

Some interest attaches to various subspaces of l^∞. We mention in particular the space of all *convergent* sequences $x = \{\xi_n\}$. We denote this space by (c).

Sometimes it is convenient to denote the norm in l^p by $\|x\|_p$, and that in l^∞ by $\|x\|_\infty$. It is worth noticing that the elements of l^p form a subclass of the elements of l^q if $1 \leqslant p < q \leqslant \infty$, and that then $\|x\|_q \leqslant \|x\|_p$ if $x \in l^p$ (see Jensen's inequality in the Introduction).

The space l^p is separable if $1 \leqslant p < \infty$. To see that this is so let us first introduce some terminology. A point $x \in l^p$ will be called rational if $x = \{\xi_n\}$ and each ξ_n is rational. In the case of the complex field a number ξ_n is called rational if its real and imaginary parts are rational. A point $x = \{\xi_n\}$ will be called of *finite type* if the set of n for which $\xi_n \neq 0$ is finite. The set of all rational points of finite type is readily seen to be countable. It is also everywhere dense in l^p. For, suppose $\epsilon > 0$ and $x \in l^p$. We choose N so that $\sum_{n=N+1}^{\infty} |\xi_n|^p < \epsilon^p/2$, and then we choose a rational point of finite type, say $y = \{\eta_n\}$, such that $\eta_n = 0$ if $n > N$, and $|\xi_k - \eta_k| < (2N)^{-1/p}\epsilon$ if $k = 1, \cdots, N$. Then

$$\|x - y\|^p = \sum_{n=1}^{N} |\xi_n - \eta_n|^p + \sum_{n=N+1}^{\infty} |\xi_n|^p < N\frac{\epsilon^p}{2N} + \frac{\epsilon^p}{2} = \epsilon^p,$$

so that $\|x - y\| < \epsilon$. Thus l^p is separable.

The space l^∞ is not separable, however. For, if $\{x_n\}$ is any countable set in l^∞, with $x_n = (\xi_1^{(n)}, \xi_2^{(n)}, \cdots)$, let $x = \{\xi_n\}$ be the element of l^∞ defined by $\xi_k = \xi_k^{(k)} + 1$ if $|\xi_k^{(k)}| \leqslant 1$ and $\xi_k = 0$ if $|\xi_k^{(k)}| > 1$. Then the kth component of $x - x_k$ is $\xi_k - \xi_k^{(k)}$, and $|\xi_k - \xi_k^{(k)}| \geqslant 1$, so that $\|x - x_k\| \geqslant 1$. Thus the set $\{x_k\}$ cannot be dense in l^∞, and l^∞ is not separable.

Example 3. Let T be any nonempty set, and let $B(T)$ be the class of all bounded functions x defined on T, with either real or complex values. Then $B(T)$ becomes a normed linear space if we define

$$\|x\| = \sup_{t \in T} |x(t)|.$$

The scalar field for $B(T)$ will be the same as the field in which the function values are required to lie.

Observe that l^∞ is the special case of $B(T)$ in which T is the set of positive integers. In the special case where T is an interval $[a, b]$ of the real axis, we shall denote $B(T)$ by $B[a, b]$.

If T is an infinite set, $B(T)$ is not separable. This is shown by the same kind of argument that was used in showing the nonseparability of l^∞.

Example 4. Suppose T is a topological space. Let $C(T)$ denote the class of bounded and continuous functions x defined on T. Then $C(T)$ is a linear manifold in $B(T)$, and we consider $C(T)$ as a normed linear space on its own account, with the norm as defined in $B(T)$. If T is compact, the requirement that the values $x(t)$ be bounded is fulfilled automatically as a consequence of the assumption that x is continuous on T. This is because of Theorem 2.2–C and the fact that a compact set of real or complex numbers is bounded. It is moreover true that, if $x \in C(T)$, there is some point $t_m \in T$ such that

$$|x(t_m)| = \max_{t \in T} |x(t)| = \|x\|.$$

This is because $|x(t)|$ is continuous on T and a compact set of real numbers is closed as well as bounded.

If T is a finite closed interval on the real axis, we shall usually denote $C(T)$ by $C[a, b]$ (see Example 3, § 1.2).

Example 5. The class $\mathscr{L}^p = \mathscr{L}^p(-\infty, \infty)$ and the corresponding linear space L^p were defined in Example 7, § 1.2. Here we assume $1 \leqq p < \infty$. If $x \in \mathscr{L}^p$ we write

$$\|x\|_p = \left(\int_{-\infty}^{\infty} |x(t)|^p \, dt\right)^{1/p}. \tag{3.11–5}$$

Evidently $\|x\|_p = 0$ is equivalent to $x(t) = 0$ a.e. and $\|x\|_p = \|y\|_p$ if $x =^0 y$. Minkowski's inequality for integrals (see the Introduction) states that

$$\|x + y\|_p \leqslant \|x\|_p + \|y\|_p$$

if $x, y \in \mathscr{L}^p$. Hence, if $[x] \in L^p$ and we define $\|[x]\| = \|x\|_p$, L^p becomes a normed linear space. As stated in § 1.2, we shall usually write x in place of $[x]$. Also, we shall write $\|x\|$ instead of $\|x\|_p$ if the situation is such that the use of the index is not essential.

Instead of functions defined on $(-\infty, \infty)$, we can consider any measurable set E in n-dimensional Euclidean space. The classes $\mathscr{L}^p$ and L^p are then defined with reference to Lebesgue integrals over E, and we indicate the dependence on E, if necessary, by writing $\mathscr{L}^p(E)$ and $L^p(E)$. Still more generally, as we shall see later, it is possible and desirable to consider the case in which E is replaced by a more general measure space and the integrals are defined in terms of a more general measure.

The space $L^p(E)$ is separable. See problem 5, § 7.3.

Example 6. Let (a, b) be an interval, finite or infinite of the real axis. A measurable function x defined on (a, b), with real or complex values, is called *essentially* bounded if there is some $A \geqslant 0$ such that the set $\{t: |x(t)| > A\}$ has measure 0 [i.e., such that $|x(t)| \leqslant A$ almost everywhere on (a, b)]. If such a constant exists, there is a least one; we call this smallest possible A the *essential least upper bound* of x, and denote it by $\sup^0 |x(t)|$. The notation $\sup^0$ distinguishes $\sup^0 |x(t)|$ from the ordinary least upper bound $\sup |x(t)|$. It can of course happen that $\sup^0 |x(t)| < \sup |x(t)|$, and it can even happen that $\sup^0 |x(t)| < \infty$ but $\sup |x(t)| = \infty$. Another characterization of $\sup^0 |x(t)|$ is the following: It is the largest number B such that if $\epsilon > 0$ the set $\{t: |x(t)| > B - \epsilon\}$ has positive measure.

Let $\mathscr{L}^\infty$ denote the class of all measurable and essentially bounded functions x defined on (a, b). If x, y $\mathscr{L}^\infty$, we write $x =^0 y$ if $x(t) = y(t)$

a.e. We then define the linear space L^∞ in relation to $\mathscr{L}^\infty$ just as we defined L^p in relation to $\mathscr{L}^p$ (see Example 7, § 1.2). If $x =^0 y$, it is clear that $\sup^0 |x(t)| = \sup^0 |y(t)|$. Hence, if we define

$$\|[x]\| = \sup^0 |x(t)|,$$

L^∞ becomes a normed linear space.

The space L^∞ is not separable. For, if $a < s < b$, let x_s be the characteristic function of the interval (a, s) as a subset of (a, b). Then it is clear that $\|[x_s] - [x_t]\| = 1$ if $a < s < t < b$. Now consider the family of spheres $\|[x] - [x_s]\| < \frac{1}{2}$ with centers at $[x_s]$ in L^∞, for various values of s. No two of these spheres have points in common, and there are uncountably many of them. It is therefore clear that any set which is everywhere dense in L^∞ cannot be countable, so that L^∞ cannot be separable.

If (a, b) is a finite interval and $x \in \mathscr{L}^\infty$, then $x \in \mathscr{L}^p$ for every $p > 0$, and it can be proved that

$$\lim_{p\to\infty} \left(\int_a^b |x(t)|^p \, dt\right)^{1/p} = \sup^0 |x(t)|. \tag{3.11–6}$$

In fact, if $A < \sup^0 |x(t)|$ and E is the set where $|x(t)| > A$, we have $m(E) > 0$ and

$$A[m(E)]^{1/p} \leqslant \left(\int_a^b |x(t)|^p \, dt\right)^{1/p} \leqslant (b - a)^{1/p} \sup^0 |x(t)|.$$

Letting $p \to \infty$, we see that

$$A \leqslant \underline{\lim} \left(\int_a^b |x(t)|^p \, dt\right)^{1/p} \leqslant \overline{\lim} \left(\int_a^b |x(t)|^p \, dt\right)^{1/p} \leqslant \sup^0 |x(t)|.$$

Since A can be as near $\sup^0 |x(t)|$ as we please, this justifies (3.11–6). If the interval (a, b) is infinite, (3.11–6) is replaced by

$$\lim_{n\to\infty} \lim_{p\to\infty} \left(\int_{E_n} |x(t)|^p \, dt\right)^{1/p} = \sup^0 |x(t)|, \tag{3.11–7}$$

where E_n is a sequence of finite intervals such that $E_1 \subset E_2 \subset \cdots$, each E_n lies in (a, b), and $\bigcup_n E_n = (a, b)$.

Example 7. The linear space $BV[a, b]$ of functions of bounded variation on $[a, b]$ was defined in Example 8, § 1.2. If $x \in BV[a, b]$ and $V(x)$ denotes the total variation of $x(t)$ for $a \leqslant t \leqslant b$, we can define a norm by

$$\|x\| = |x(a)| + V(x). \tag{3.11–8}$$

The space $BV[a, b]$ is not separable. This may be shown by the same kind of argument that was used in showing the nonseparability of L^∞.

Many interesting spaces can be formed from classes of analytic functions of a complex variable. Let $\mathfrak{A}$ denote the class of all functions $f(z)$ which are defined, single valued, and analytic in the unit circle $|z| < 1$ of the complex plane. This class is a complex linear space. We shall in the next two examples describe some subspaces of $\mathfrak{A}$ which become normed linear spaces with appropriately defined norms.

Example 8. Suppose $0 < p < \infty$. For any $f \in \mathfrak{A}$ and $0 \leqslant r < 1$ let

$$\mathfrak{M}_p[f; r] = \left(\frac{1}{2\pi}\int_0^{2\pi} |f(re^{i\theta})|^p \, d\theta\right)^{1/p}. \tag{3.11–9}$$

The class H^p is by definition composed of those $f \in \mathfrak{A}$ such that $\sup\limits_{0 \leqslant r < 1} \mathfrak{M}_p[f; r] < \infty$. If $1 \leqslant p$ it is clear by Minkowski's inequality that $f + g$ is in H^p if f and g are in H^p. If $0 < p < 1$, this same conclusion follows from the following inequality [Hardy, Littlewood, and Pólya, **1**, formula (6.13–6)]:

$$\int_0^{2\pi} |f(re^{i\theta}) + g(re^{i\theta})|^p \, d\theta \leqslant \int_0^{2\pi} |f(re^{i\theta})|^p \, d\theta + \int_0^{2\pi} |g(re^{i\theta})|^p d\theta. \tag{3.11–10}$$

It is thus seen that H^p is a linear space. If $1 \leqslant p$ we define

$$\|f\| = \sup_{0 \leqslant r < 1} \mathfrak{M}_p[f; r]. \tag{3.11–11}$$

This has the properties of a norm; thus H^p is a normed linear space if $1 \leqslant p$. If $0 < p < 1$, (3.11–11) does not define a norm, for the triangularity condition $\|f + g\| \leqslant \|f\| + \|g\|$ is not always satisfied.

It is natural to define H^∞ as the class of those $f \in \mathfrak{A}$ such that the values $f(z)$ are bounded. We define a norm on H^∞ by

$$\|f\| = \sup_{|z|<1} |f(z)|.$$

That H^∞ is a natural notation for this space is seen from the fact that, if $f \in \mathfrak{A}$ and $0 \leqslant r < 1$, then

$$\max_{|z|=r} |f(z)| = \lim_{p\to\infty} \mathfrak{M}_p[f; r],$$

while, if $f \in H^\infty$, then

$$\|f\| = \sup_{0 \leqslant r < 1} \left\{\max_{|z|=r} |f(z)|\right\}.$$

For complete uniformity of notation we define

$$\mathfrak{M}_\infty[f; r] = \max_{|z|=r} |f(z)|.$$

Then (3.11–11) is the definition of the norm in H^p for $1 \leqslant p \leqslant \infty$, and H^p consists of those $f \in \mathfrak{A}$ such that $\|f\| < \infty$.

The space H^p is separable if $1 \leqslant p < \infty$. This may be shown in various ways. If one makes use of the fact that L^p is separable, the separability of H^p follows at once from the following facts: If $f \in H^p$, then $\lim\limits_{r \to 1} f(re^{i\theta})$ exists for almost all values of θ, thus defining a function which we denote by $f(e^{i\theta})$. Moreover, the latter function belongs to $L^p(0, 2\pi)$, and

$$\|f\| = \left(\frac{1}{2\pi}\int_0^{2\pi} |f(e^{i\theta})|^p \, d\theta\right)^{1/p}$$

For these facts about H^p, we refer the reader to F. Riesz, **2**, and A. Zygmund, **1**, pages 158–162. Thus H^p is in isometric correspondence with a subset of L^p. Since any subset of a separable metric space is itself separable, we see that H^p is separable.

The space H^∞ is not separable, however. To show this, it suffices to show that the set of all $f \in H^\infty$ such that $\|f\| = 1$ is not separable. Let $\{f_n\}$ be any sequence from this set. Then $\mathfrak{M}_\infty[f_n; r] \to 1$ as $r \to 1$, for each n. Choose $r_n > 0$ so that $r_n \leqslant r < 1$ implies $\mathfrak{M}_\infty[f_n; r] \geqslant {}^1/_2$. Let $\rho_n = \max\{r_n, 1 - 1/n^2\}$, and choose points α_n so that $|\alpha_n| = \rho_n$ and $|f_n(\alpha_n)| = \mathfrak{M}_\infty[f_n; \rho_n]$. Thus $1 - 1/n^2 \leqslant |\alpha_n| < 1$ and $|f_n(\alpha_n)| \geqslant {}^1/_2$. Now define $f(z)$ by the Blaschke product

$$f(z) = \prod_{k=1}^{\infty} |\alpha_k| \frac{1 - (z/\alpha_k)}{1 - \bar{\alpha}_k z}.$$

Then $f \in H^\infty$, $f(\alpha_k) = 0$, and $\|f\| = 1$, for it is known that $|f(e^{i\theta})| = 1$ for almost all values of θ. (See pages 160–161 and pages 163–164 of Zygmund, **1**, or F. Riesz, **2**.) Finally, $\|f_n - f\| \geqq |f_n(\alpha_n) - f(\alpha_n)| \geqq {}^1/_2$. Thus H^∞ cannot be separable.

Example 9. Let CA denote the class of functions f which are defined (single valued) and continuous when $|z| \leqslant 1$ and analytic when $|z| < 1$. Evidently CA is a linear subclass of H^∞ and is a normed linear space with the norm defined as in H^∞. Because of the maximum modulus theorem and the fact that f is continuous, it is clear that for $f \in CA$

$$\|f\| = \max_{|z|=1} |f(z)|. \tag{3.11–12}$$

The space CA is separable. In fact, the linear combinations of the functions $1, z, z^2, \cdots, z^n, \cdots$, with complex rational coefficients, form a set everywhere dense in CA. This can be seen, for instance, from Fejér's theorem on the $(C, 1)$ summability of Fourier series. If

$$f(z) = \sum_{0}^{\infty} a_n z^n$$

is in CA, the Fourier series of $f(e^{i\theta})$ is

$$\sum_{0}^{\infty} a_n e^{in\theta},$$

and this series is uniformly $(C, 1)$ summable to $f(e^{i\theta})$ (Zygmund, **1**, page 45). This implies that

$$\sum_{k=0}^{n-1} (1 - k/n) a_k z^k$$

converges to $f(z)$ uniformly in $|z| \leqslant 1$ as $n \to \infty$.

The spaces H^p with $1 \leqslant p \leqslant \infty$ and the space CA have been investigated in connection with more general studies of Banach spaces composed of functions which are analytic in the unit circle. See Taylor, **4**.

PROBLEMS

1. The complex space $C(T)$ is separable if and only if the corresponding *real* space is separable.

2. The real space $C[a, b]$ is separable if $[a, b]$ is a finite closed interval of the real axis. One way of seeing this may be outlined as follows: For each positive integer n and each set of $n + 1$ rational numbers $r_0, r_1, \cdots, r_n$, divide $[a, b]$ into $2n$ equal parts by points $t_0, t_1, \cdots, t_{2n}$ (in order from a to b); then define a function x by setting $x(t) = r_k$ if $t_{2k} \leqslant t \leqslant t_{2k+1}$, $k = 0, 1, \cdots, n - 1$, $x(t_{2n}) = r_n$, making $x(t)$ linear in each of the remaining subintervals in such a way that it is continuous at all points. The set of such functions is everywhere dense in $C[a, b]$.

3. The separability of $C[a, b]$ may also be established by relying on the known theorem of Weierstrass which states (in effect) that those elements of $C[a, b]$ which are polynomials in t are dense in $C[a, b]$. One then shows that the polynomials with rational coefficients are also dense in $C[a, b]$.

3.12 Finite-Dimensional Normed Linear Spaces

We saw in Theorem 1.4–A that two linear spaces of the same finite dimension n over the same scalar field are isomorphic. Now we shall see that, if each of the two spaces has a norm, the spaces are topologically isomorphic according to the definition of § 3.1.

Theorem 3.12–A. *Let* X_1 *and* X_2 *be two normed linear spaces of the same finite dimension n, with the same scalar field. Then* X_1 *and* X_2 *are topologically isomorphic.*

PROOF. The case $n = 0$ is trivial, and we assume $n \geqslant 1$. It will suffice to prove that, if X is an n-dimensional normed linear space, it is topologically isomorphic to $l^1(n)$, for the relation of topological isomorphism is transitive (as well as being reflexive and symmetric). Suppose that $x_1, \cdots, x_n$ is a basis for X. If $x = \xi_1 x_1 + \cdots + \xi_n x_n$ is the representation of an arbitrary $x \in X$, we know from § 1.4 that the correspondence $x \leftrightarrow (\xi_1, \cdots, \xi_n)$ defines an isomorphism of X and $l^1(n)$. By Theorem 3.1–C, all we have to prove is that there exists positive constants m and M such that

$$\|\xi_1 x_1 + \cdots + \xi_n x_n\| \leqslant M(|\xi_1| + \cdots + |\xi_n|) \tag{3.12–1}$$

and

$$m(|\xi_1| + \cdots + |\xi_n|) \leqslant \|\xi_1 x_1 + \cdots + \xi_n x_n\| \tag{3.12–2}$$

for all possible sets of values $(\xi_1, \cdots, \xi_n)$. Now (3.12–1) is evidently true if we choose for M the maximum of $\|x_1\|, \cdots, \|x_n\|$. To prove (3.12–2) it suffices to show that $m \leqslant \|\xi_1 x_1 + \cdots + \xi_n x_n\|$ if $|\xi_1| + \cdots + |\xi_n| = 1$, for (3.12–2) certainly holds if $\xi_1 = \cdots = \xi_n = 0$, and, if $c = |\xi_1| + \cdots + |\xi_n| > 0$, we can define $\eta_i = c^{-1}\xi_i$, whence $|\eta_1| + \cdots + |\eta_n| = 1$, and $\|\eta_1 x_1 + \cdots + \eta_n x_n\| = c^{-1}\|\xi_1 x_1 + \cdots + \xi_n x_n\|$, so that (3.12–2) holds if $m \leqslant \|\eta_1 x_1 + \cdots + \eta_n x_n\|$.

Now let $f(\xi_1, \cdots, \xi_n) = \|\xi_1 x_1 + \cdots + \xi_n x_n\|$. The function f is continuous on $l^1(n)$, as a consequence of (3.12–1). In fact, since $|\,\|x\| - \|y\|\,| \leqslant \|x - y\|$, we see that

$$\begin{aligned} |f(\xi_1, \cdots, \xi_n) - f(\eta_1, \cdots, \eta_n)| &\leqslant \|(\xi_1 - \eta_1)x_1 + \cdots + (\xi_n - \eta_n)x_n\| \\ &\leqslant M(|\xi_1 - \eta_1| + \cdots + |\xi_n - \eta_n|), \end{aligned}$$

whence the continuity of f is clear. Now the surface of the unit sphere in $l^1(n)$ (the set S for which $|\xi_1| + \cdots + |\xi_n| = 1$) is compact, as we see, for example, by the use of Theorem 2.4–H and the Bolzano-Weierstrass theorem for scalars. Hence f, being continuous on S, attains a minimum value $m \geqslant 0$ on S. But $m > 0$, for $m = 0$ would imply that $x_1, \cdots, x_n$ are linearly dependent, contrary to the fact that they form a basis for X. The proof of (3.12–2) and of Theorem 3.12–A is now complete.

It is clear from Theorem 3.1–C that, if X and Y are topologically isomorphic normed linear spaces and if one of them is complete (as a metric space), the other is also complete. Now $l^1(n)$ is evidently complete (as a consequence of the completeness of $l^1(1)$, the real number field). Thus we have:

Theorem 3.12–B. *A finite-dimensional normed linear space is complete.*

As a corollary of this we have:

Theorem 3.12–C. *If X is a normed linear space, any finite-dimensional subspace of X is necessarily closed.*

Another important result is the following:

Theorem 3.12–D. *If X is a finite-dimensional normed linear space, each closed and bounded set in X is compact.*

PROOF. This proposition is true (by classical analysis) for the particular finite-dimensional space $l^1(n)$. It then follows, by virtue of Theorem 3.12–A, that the theorem is true for any finite-dimensional space X, for the properties of being bounded and closed are transferred from a set S to its image S_1 in $l^1(n)$ by the topological isomorphism, and the compactness is then carried back from S_1 to S.

The converse of Theorem 3.12–D is also true. Before proving the converse, we consider a general theorem due to F. Riesz which is useful in many arguments.

Theorem 3.12–E (Riesz's Lemma). *Suppose X is a normed linear space. Let X_0 be a subspace of X such that X_0 is closed and a proper subset of X. Then for each θ such that $0 < \theta < 1$ there exists a vector $x_\theta \in X$ such that $\|x_\theta\| = 1$ and $\|x - x_\theta\| \geqslant \theta$ if $x \in X_0$.*

PROOF. Select any $x_1 \in X - X_0$ and let

$$d = \inf_{x \in X_0} \|x - x_1\|.$$

Since X_0 is closed, it follows that $d > 0$. There exists $x_0 \in X_0$ such that $\|x_0 - x_1\| \leqslant \theta^{-1}d$ (because $\theta^{-1}d > d$). Let $x_\theta = h(x_1 - x_0)$, where $h = \|x_1 - x_0\|^{-1}$. Then $\|x_\theta\| = 1$. If $x \in X_0$, then $h^{-1}x + x_0 \in X_0$ also, and so

$$\|x - x_\theta\| = \|x - hx_1 + hx_0\| = h\|(h^{-1}x + x_0) - x_1\| \geqslant hd.$$

But $hd = \|x_1 - x_0\|^{-1}d \geqslant \theta$, by the way in which x_0 was chosen. Thus $\|x - x_\theta\| \geqslant \theta$ if $x \in X_0$, and the proof is complete.

We can restate Riesz's lemma as follows: *If X_0 is a closed and proper subspace of X, there exist on the surface of the unit sphere in X points whose distance from X_0 is as near 1 as we please.* This is the most that can be said in general however. It need not be true that there are points on the unit sphere whose distance from X_0 is *equal* to 1.

Example. Let X be that subspace of the real space $C[0, 1]$ (see § 3.11, Example 4) consisting of all continuous functions x on $[0, 1]$ such that $x(0) = 0$. For X_0 we take the subspace of all $x \in X$ such that $\int_0^1 x(t)\,dt = 0$. Now suppose that $x_1 \in X$, $\|x_1\| = 1$, and $\|x_1 - x\| \geqslant 1$ if $x \in X_0$. Corresponding to each $y \in X - X_0$ let

$$c = \frac{\int_0^1 x_1(t)\,dt}{\int_0^1 y(t)\,dt}.$$

Then $x_1 - cy \in X_0$, and so $1 \leqslant \|x_1 - (x_1 - cy)\| = |c|\,\|y\|$, or

$$\left|\int_0^1 y(t)\,dt\right| \leqslant \left|\int_0^1 x_1(t)\,dt\right| \|y\|$$

for each $y \in X$. Now, we can make $\left|\int_0^1 y(t)\,dt\right|$ as close to 1 as we please while maintaining $\|y\| = 1$ (e.g., let $y_n(t) = t^{1/n}$ and let $n \to \infty$). Thus we see that

$$1 \leqslant \left|\int_0^1 x_1(t)\,dt\right|.$$

But, since $\|x_1\| = \max_{0 \leqslant t \leqslant 1} |x_1(t)| = 1$ and $x_1(0) = 0$, the continuity of x_1 shows that we must have

$$\left|\int_0^1 x_1(t)\,dt\right| < 1,$$

and thus we have a contradiction. Therefore, with X_0 and X as here given, there is no point on the surface of the unit sphere in X at unit distance from X_0.

We come now to the converse of Theorem 3.12–D.

Theorem 3.12–F. *Let X be a normed linear space, and suppose the surface S of the unit sphere in X is compact. Then X is finite dimensional.*

Suppose that X is not finite dimensional. Choose $x_1 \in S$, and let X_1 be the subspace generated by x_1. Then X_1 is a proper subspace of X, and it is closed (by Theorem 3.12–C). Hence, by Riesz's lemma, there exists $x_2 \in S$ such that $\|x_2 - x_1\| \geqslant 1/2$. Let X_2 be the (closed and proper) subspace of X generated by x_1, x_2; then, there must exist $x_3 \in S$ such that $\|x_3 - x\| \geqslant 1/2$ if $x \in X_2$. Proceeding by induction, we obtain an infinite sequence $\{x_n\}$ of elements of S such that $\|x_n - x_m\| \geqslant 1/2$ if $m \neq n$. This sequence can have no convergent subsequence. This contradicts Theorem 2.4–H, since S is compact. Thus X must be finite dimensional.

Another proof, using the definition of compactness directly, may be given as follows: The family of all open spheres of radius $1/2$ with centers on S is an open covering of S. Since S is compact, there must exist a finite number of points $x_1, \cdots, x_n$ on S such that S is covered by the set of open spheres of radius $1/2$ with centers at $x_1, \cdots, x_n$. Let M be the finite dimensional, and therefore closed, subspace of X generated by $x_1, \cdots, x_n$. Then M must be all of X. For, if not, by Riesz's lemma there exists a point $x_0 \in S$ whose distance from M is greater than $1/2$, and this point x_0 cannot be in any of the spheres which cover S. Since $M = X$, X is finite dimensional.

PROBLEM

If X and Y are normed linear spaces, if X is finite dimensional, and if T is a linear operator on X into Y, T is necessarily continuous.

3.13 Banach Spaces

If a normed linear space is complete, it is called a *Banach space*. While many propositions about normed linear spaces do not require the hypothesis of completeness, a number of theorems of critical importance do depend upon completeness. In particular, there are some important theorems in the theory of linear operators which make use of the theorem of Baire (Theorem 2.41–C), and this is made possible by the assumption of completeness for the normed linear spaces under consideration. In some work it is necessary to construct elements of a normed linear space by means of infinite series or integrals, and completeness is then needed to insure the existence of limits.

If a normed linear space is not complete, we may complete it as a metric space by the process described in the proof of Theorem 2.41–A. It is important to know that we can complete X, not merely as a metric space, but also as a normed linear space. That is, using the notation of Theorem 2.41–A, where X is the normed linear space and Y is its completion as a metric space, the space Y can be made into a linear space in such a way that the subset Y_0 (which is isometric with X and dense in Y) is congruent to X as a normed linear space.

Suppose $y, v \in Y$, and let $\{x_n\}$, $\{u_n\}$ be Cauchy sequences from X which are in the equivalence classes y, v respectively. Then $\{x_n + u_n\}$ is a Cauchy sequence, since

$$\|(x_n + u_n) - (x_m + u_m)\| \leqslant \|x_n - x_m\| + \|u_n - u_m\|.$$

Moreover, if $\{a_n\}$ and $\{b_n\}$ are Cauchy sequences equivalent to $\{x_n\}$ and $\{u_n\}$ respectively, then $\{a_n + b_n\}$ is equivalent to $\{x_n + u_n\}$, because

$$\|(x_n + u_n) - (a_n + b_n)\| \leqslant \|x_n - a_n\| + \|u_n - b_n\|.$$

Thus we can define $y + v$ as that equivalence class which contains $\{x_n + u_n\}$, and the definition depends only on y, v, not on the particular choice of $\{x_n\}$, $\{u_n\}$. Likewise we define αy as the equivalence class which contains $\{\alpha x_n\}$, and we define $\|y\| = \lim\limits_{n\to\infty} \|x_n\|$. The zero element of Y is the unique equivalence class all of whose members $\{x_n\}$ are such that $x_n \to 0$. It is a routine matter to verify that Y is a normed linear space and that Y_0 (defined in the proof of Theorem 2.41–A) is a linear manifold in Y congruent to X. Finally, the metric defined in Y by the norm is the same as the metric defined in Y by the completion process of Theorem 2.41–A when the metric of X is defined by the norm in X.

It is convenient to adopt a standard notation and terminology regarding the completion of a space X. It is evident that we can adjoin certain "ideal" elements to X so as to obtain a complete space in which X is everywhere dense. This complete space will be called the completion of X, and denoted by $\hat{X}$. The important feature of this arrangement is that X is actually a subspace of $\hat{X}$, whereas in the original process of completing a metric space, we obtain a complete space Y whose elements are equivalence classes of Cauchy sequences from X, so that X is not actually a subset of Y. A space X is complete if and only if $X = \hat{X}$.

There is a particular kind of extension theorem for linear operators which we shall need later and which is convenient to discuss here.

Theorem 3.13–A. *Let X and Y be normed linear spaces, and let T be a continuous linear operator on X into Y. Then there is a uniquely determined continuous linear operator $\hat{T}$ on $\hat{X}$ into $\hat{Y}$ such that $\hat{T}x = Tx$ if $x \in X$. The relation $\|\hat{T}\| = \|T\|$ is valid.*

PROOF. To define $\hat{T}$ we suppose $\hat{x} \in \hat{X}$, and select a sequence $\{x_n\}$ from X such that $x_n \to \hat{x}$. Then $\{x_n\}$ is a Cauchy sequence, and

$$\|Tx_n - Tx_m\| = \|T(x_n - x_m)\| \leqslant \|T\|\,\|x_n - x_m\|,$$

so that $\{Tx_n\}$ is a Cauchy sequence in Y. Consequently $Tx_n \to \hat{y}$, where $\hat{y}$ is some element of $\hat{Y}$. It is readily proved that $\hat{y}$ depends only on $\hat{x}$ and T, not on the particular sequence $\{x_n\}$. We define $\hat{T}\hat{x} = \hat{y}$. It is a simple matter to verify that $\hat{T}x = Tx$ if $x \in X$ and that $\hat{T}$ is linear. We see that $\|x_n\| \to \|\hat{x}\|$, $\|Tx_n\| \leqslant \|T\|\,\|x_n\|$, and hence $\|\hat{T}\hat{x}\| \leqslant \|T\|\,\|\hat{x}\|$. Thus $\hat{T}$ is continuous and $\|\hat{T}\| \leqslant \|T\|$ (see § 3.1). On the other hand, if $x \in X$, we have $\|Tx\| = \|\hat{T}x\| \leqslant \|\hat{T}\|\,\|x\|$, so that $\|T\| \leqslant \|\hat{T}\|$. Hence $\|\hat{T}\| = \|T\|$. The uniqueness assertion about $\hat{T}$ is easily justified by using the fact that X is dense in $\hat{X}$.

Theorem 3.13–B. *If X is a Banach space and X_0 is a closed linear manifold in X, then X_0, considered as a normed linear space by itself, is a Banach space.*

PROOF. This is a special case of Theorem 2.41–E.

It is useful to note that a Cauchy sequence $\{x_n\}$ in a metric space is bounded, irrespective of whether or not the space is complete. Hence, if $\{x_n\}$ is a Cauchy sequence in a normed linear space, the sequence of norms $\|x_n\|$ is bounded.

One simple but important property of Banach spaces is shown in the following theorem:

Theorem 3.13–C. *Let X be a Banach space, and let $\{x_n\}$ be a sequence of elements of X such that $\sum_1^\infty \|x_n\| < \infty$. Then, the infinite series $\sum_1^\infty x_n$ is convergent and defines an element of X.*

PROOF. Let $y_n = x_1 + \cdots + x_n$. Then $\{y_n\}$ is a Cauchy sequence, for, if $m < n$, $\|y_n - y_m\| \leqslant \|x_{m+1}\| + \cdots + \|x_n\|$, and we can employ the Cauchy criterion on the series $\sum \|x_n\|$. Since X is complete, it follows that $\lim_{n\to\infty} y_n$ exists.

We turn to a discussion of completeness for the examples of normed linear spaces given in § 3.11. We already know (Theorem 3.12–B) that every finite-dimensional normed linear space is complete.

The space l^p, where $p \geqslant 1$, is complete. Let $\{x_n\}$ be a Cauchy sequence in l^p, with $x_n = (\xi_1^{(n)}, \xi_2^{(n)}, \cdots)$. For each fixed k, $\{\xi_k^{(n)}\}$ is a Cauchy sequence, because

$$|\xi_k^{(n)} - \xi_k^{(m)}| \leqslant \left(\sum_{i=1}^\infty |\xi_i^{(n)} - \xi_i^{(m)}|^p\right)^{1/p} = \|x_n - x_m\|.$$

Let $\xi_k = \lim_{n\to\infty} \xi_k^{(n)}$. We shall first prove that the sequence $\{\xi_k\}$ is an element of l^p. We know that $\|x_n\|$ is bounded, say $\|x_n\| \leqslant M$. Now, for any k,

$$\left(\sum_{i=1}^k |\xi_i^{(n)}|^p\right)^{1/p} \leqslant \|x_n\| \leqslant M.$$

Letting $n \to \infty$, we obtain

$$\left(\sum_{i=1}^k |\xi_i|^p\right)^{1/p} \leqslant M.$$

Since k is arbitrary, this shows that $\{\xi_k\} \in l^p$ and that its norm does not exceed M. Let $x = \{\xi_k\}$. It remains to prove that $\|x_n - x\| \to 0$. Suppose $\epsilon > 0$. Then there exists an integer N such that $\|x_n - x_m\| < \epsilon$ if $N \leqslant m$ and $N \leqslant n$. Therefore, for any k,

$$\left(\sum_{i=1}^{k} |\xi_i^{(n)} - \xi_i^{(m)}|^p\right)^{1/p} \leqslant \|x_n - x_m\| < \epsilon$$

if $N \leqslant m$ and $N \leqslant n$. Keeping k and n fixed, let $m \to \infty$. This gives

$$\left(\sum_{i=1}^{k} |\xi_i^{(n)} - \xi_i|^p\right)^{1/p} \leqslant \epsilon$$

if $N \leqslant n$. Since this is true for all k, we can let $k \to \infty$, and we obtain the result that $\|x_n - x\| \leqslant \epsilon$ if $N \leqslant n$. This finishes the proof that l^p is complete.

The form of the foregoing argument is such that it can be adapted to proving the completeness of a number of spaces. To avoid much repetition of essentially the same argument we shall formulate a general theorem which embodies the principle of the argument.

Theorem 3.13–D. *Let Y be a linear space. Let S be a certain class of sequences $\{y_n\}$ of elements of Y, and let $\mathscr{F}$ be a certain family of real-valued functions defined on Y. We assume that S and $\mathscr{F}$ satisfy the following conditions:*

1. *With each $\{y_n\} \in S$ is associated a certain element $y \in Y$. This association is indicated by writing $y_n \to y$.*
2. *If $\{y_n\} \in S$, with $y_n \to y$, then for each fixed k the sequence $\{z_n\}$ with $z_n = y_k - y_n$ is in S,* and *$z_n \to y_k - y$.*
3. *If $\{y_n\} \in S$, with $y_n \to y$, then $f(y_n) \to f(y)$ for each $f \in \mathscr{F}$.*
4. *The class X of $x \in Y$ such that $\sup_{f\in\mathscr{F}} f(x) < +\infty$ is a linear manifold in Y and the function $\|x\|$ on X defined by*

$$\|x\| = \sup_{f\in\mathscr{F}} f(x)$$

is a norm on X.

5. *If $\{x_n\}$ is a Cauchy sequence in X [relative to the metric topology defined by the norm of* (4)], *then $\{x_n\} \in S$.*

With these assumptions, X is a Banach space (i.e., it is complete).

PROOF. Suppose that $\{x_n\}$ is a Cauchy sequence in X. Then there is some M such that $\|x_n\| \leqslant M$ for all values of n. By (1) and (5) there is some $x \in Y$ such that $x_n \to x$, and by (3) it follows that $f(x_n) \to f(x)$ for each $f \in \mathscr{F}$. Now $f(x_n) \leqslant \|x_n\| \leqslant M$ for each $f \in \mathscr{F}$, by the definition of the norm. Letting $n \to \infty$, we see that $f(x) \leqslant M$. Since this is true for every $f \in \mathscr{F}$, we see by (4) that $x \in X$ and $\|x\| \leqslant M$. Now, if $\epsilon > 0$, there exists an integer N such that $\|x_n - x_m\| < \epsilon$ if $N \leqslant n$ and $N \leqslant m$. Therefore, for each $f \in \mathscr{F}$, $f(x_n - x_m) < \epsilon$ under the same conditions on m and n. Letting $m \to \infty$, we see by (2) and (3) that $f(x_n - x_m) \to f(x_n - x)$, and so $f(x_n - x) \leqslant \epsilon$. This is true for each $f \in \mathscr{F}$ and each $n \geqslant N$. Therefore $\|x_n - x\| \leqslant \epsilon$ if $N \leqslant n$. This finishes the proof that X is complete.

To see the application of this theorem to l^p let Y be the linear space of *all* sequences $y = \{\eta_i\}$, where $\eta_1, \eta_2, \cdots$ are scalars. Let S be the class of all sequences $\{y_n\}$ in Y (where $y_n = \{\eta_i^{(n)}\}$) such that $\lim\limits_{n\to\infty} \eta_i^{(n)} = \eta_i$ exists for $i = 1, 2, \cdots$, and let the associated $y \in Y$ be $\{\eta_i\}$. Let $\mathscr{F}$ be the countable family of functions $f_1, f_2, \cdots$, with f_k defined by

$$f_k(y) = \left(\sum_{i=1}^{k} |\eta_i|^p\right)^{1/p}, \qquad y \in Y.$$

Then Theorem 3.13–D can be applied with $X = l^p$. Our proof that l^p is complete follows exactly the same lines as the proof of Theorem 3.13–D.

It is easy to give examples of incomplete normed linear spaces. Consider, for example, the subset X of l^p $(1 \leqslant p < \infty)$ consisting of sequences $x = \{\xi_n\}$ such that the set of n for which $\xi_n \neq 0$ is finite. This subset of l^p is a linear manifold and can be regarded as a space by itself. It is not complete. For, let x be an element of l^p that is not in X. If $x = \{\xi_n\}$, let $x_n = (\xi_1, \xi_2, \cdots, \xi_n, 0, 0, \cdots)$. Then $x_n \in X$, and $\{x_n\}$ is a Cauchy sequence in X. But the sequence has no limit in X, for, if it did, say $x_n \to y$, we should have $y = x$, since it is readily seen that $x_n \to x$, and the limit of a convergent sequence is unique. But $y \in X$ and $x \notin X$. Thus we have a contradiction, and we have proved that X is not complete.

The essential principle in the foregoing argument shows us the truth of the following statement: If X is a linear manifold in a normed linear space Y such that X is dense in Y but a proper subset of Y, then X, considered as a normed linear space by itself, is not complete.

PROBLEMS

1. Theorem 3.13–D may be used to prove that a number of spaces are complete. Indications of procedure are given below. Details are left to the reader.

a. B(T) is complete (see Example 3, § 3.11). Let Y be the linear space of *all* scalar-valued functions y defined on T. Let S be the class of sequences $\{y_n\}$

from Y such that $\lim_{n\to\infty} y_n(t) = y(t)$ exists for each $t \in T$; then write $y_n \to y$. Let $\mathscr{F}$ be the family of functions f_t, $t \in T$, defined by $f_t(y) = y(t)$.

b. $BV[a, b]$ is complete (see Example 7, § 3.11). Choose Y and S as in (*a*), with $T = [a, b]$. For $\mathscr{F}$ take the class of all function f_Δ, defined by

$$f_\Delta(y) = |y(a)| + \sum_{i=1}^{n} |y(t_i) - y(t_{i-1})|,$$

where Δ is the partition of $[a, b]$ by the points $t_0, t_1, \cdots, t_n$ $(a = t_0 < t_1 < \cdots < t_n = b)$; $\mathscr{F}$ is generated by taking all possible Δ's.

c. H^p is complete, $1 \leqslant p \leqslant \infty$ (see Example 8, § 3.11). Let Y be the class $\mathfrak{A}$ (defined previously) of all functions y of the complex variable t, defined and analytic when $|t| < 1$. Let S be the class of sequences $\{y_n\}$ from Y such that $y_n(t)$ is convergent to a limit $y(t)$ as $n \to \infty$, uniformly with respect to t in every compact set lying in the open unit circle $|t| < 1$. For $\mathscr{F}$ we take the family of functions f_r defined by

$$f_r(y) = \mathfrak{M}_p[y; r], \qquad 0 \leqslant r < 1.$$

The meaning of $\mathfrak{M}_p$ is explained in § 3.11; we are now using y where f was used in § 3.11. The X of Theorem 3.13–D is H^p in this case. For $x \in X$ it may be proved that

$$|x(t)| \leqslant \frac{\|x\|}{1 - |t|}$$

if $|t| < 1$, and from this it follows that condition 5 in Theorem 3.13–D is satisfied. The proof of the inequality runs as follows: Let

$$x(t) = \sum_{n=0}^{\infty} \xi_n t^n$$

$$\xi_n = \frac{1}{2\pi r^n} \int_0^{2\pi} x(re^{i\theta}) e^{-in\theta}\, d\theta, \qquad 0 < r < 1.$$

Then

$$r^n |\xi_n| \leqslant \mathfrak{M}_p[x; r] \leqslant \|x\|,$$

whence $|\xi_n| \leqslant \|x\|$ and

$$|x(t)| \leqslant \sum_{n=0}^{\infty} \|x\|\, |t|^n = \frac{\|x\|}{1 - |t|}.$$

If $p = \infty$ we have the stronger inequality

$$|x(t)| \leqslant \|x\|.$$

d. The space $C(T)$ is complete (Example 4, § 3.11). It suffices to show that $C(T)$ is closed in $B(T)$. This follows from the fact that in $B(T)$ $\|x_n - x\| \to 0$ signifies uniform convergence of $x_n(t)$ to $x(t)$.

e. The space CA is complete (Example 9, § 3.11). Argument as in (*d*), with H^∞ in place of $B(T)$.

2. *The space $L^p(E)$ is complete* (Example 5, § 3.11). Here $1 \leqslant p < \infty$ and E is a measurable set in Euclidean space of n-dimensions. This fundamental theorem in the theory of Lebesgue integration goes back to work of Fischer and F. Riesz; see, e.g., Titchmarsh, **1**, § 12.5. For the result in a more abstract setting see § 7.1.

3. *The space* L^∞ *is complete* (Example 6, § 3.11). The argument is left to the reader.

3.14 Quotient Spaces

Let X be a linear space, and M a linear manifold in X. Two elements $x_1, x_2 \in X$ are said to be *equivalent modulo M* if $x_1 - x_2 \in M$. We write $x_1 \equiv x_2 \pmod{M}$. It is clear that this kind of equivalence has the usual properties of an equivalence relation, namely reflexivity, symmetry, and transitivity. Hence X is divided into mutually disjoint equivalence classes, two elements being in the same equivalence class if and only if they are equivalent modulo M. The set of all these equivalence classes is denoted by X/M. We shall explain how to define addition and multiplication by scalars in X/M so that it becomes a linear space. Let $[x]$ denote the equivalence class which contains the element x; thus $[x_1] = [x_2]$ if and only if $x_1 \equiv x_2 \pmod{M}$. We make the definition $[x] + [y] = [x + y]$. To show that $[x] + [y]$ is unambiguously defined we have to show that $[u + v] = [x + y]$ if $[u] = [x]$ and $[v] = [y]$. This is at once evident, however, because $(x + y) - (u + v) = (x - u) + (y - v)$ and M is a linear manifold. We also define $\alpha[x] = [\alpha x]$, observing that $[\alpha x] = [\alpha y]$ if $[x] = [y]$. It is a routine matter to verify that X/M becomes a linear space as a result of these definitions. The zero element of X/M is $[0]$, which is the same as M.

Definition. The linear space X/M, as described in the foregoing paragraph, is called the *quotient space* of X modulo M. The mapping ϕ of X onto X/M defined by $\phi(x) = [x]$ is called the *canonical mapping* of X onto X/M.

To get an intuitive geometric appreciation of the definition of X/M, consider the case in which X is three dimensional and the points of X are represented in a three-dimensional rectangular-coordinate system, the point $x = (\xi_1, \xi_2, \xi_3)$ having ξ_1, ξ_2, ξ_3 as its coordinates. If M is the linear manifold of points $(\xi_1, 0, 0)$ (i.e., the ξ_1-axis), the elements of X/M are the straight lines parallel to the ξ_1-axis. Each such line is uniquely determined by the point $(0, \xi_2, \xi_3)$ in which it intersects the plane $\xi_1 = 0$, and it is readily seen that X/M is isomorphic to the linear manifold of points $(0, \xi_2, \xi_3)$. Again, if M is the linear manifold of points $(\xi_1, \xi_2, 0)$, the elements of X/M are the planes parallel to the plane $\xi_3 = 0$. Each such plane is uniquely determined by the point $(0, 0, \xi_3)$ in which it is pierced by the ξ_3-axis, and X/M is isomorphic to the one-dimensional linear manifold of points $(0, 0, \xi_3)$. The isomorphism between X/M and a certain subspace of X, which is apparent in these particular cases, persists in general, but we shall not pursue this aspect of the situation further.

If X is a normed linear space and the linear manifold M is closed, it is possible to define a norm on X/M.

Theorem 3.14–A. *Let M be a closed linear manifold in the normed linear space X. For each element $[x] \in X/M$ we define*

$$\|[x]\| = \inf_{y\in[x]} \|y\|. \tag{3.14–1}$$

Then $\|[x]\|$ is a norm on X/M. If X is a Banach space, so is X/M.

PROOF. We have

$$\|[x_1] + [x_2]\| = \|[x_1 + x_2]\| = \inf_{y\in[x_1+x_2]} \|y\| = \inf_{u\in[x_1],\, v\in[x_2]} \|u + v\|$$

$$\leqslant \inf_{u\in[x_1]} \|u\| + \inf_{v\in[x_2]} \|v\| = \|[x_1]\| + \|[x_2]\|.$$

Also

$$\|\alpha[x]\| = \|[\alpha x]\| = \inf_{y\in[\alpha x]} \|y\| = \inf_{u\in[x]} \|\alpha u\| = |\alpha| \inf_{u\in[x]} \|u\| = |\alpha|\, \|[x]\|.$$

To show that $\|[x]\|$ is a norm it remains only to show that $[x]$ is the zero element of X/M (i.e., that $x \in M$) if $\|[x]\| = 0$. Now M is a closed set, and from this it follows that each of the equivalence classes $[x]$ is closed. In fact, $[x]$ is homeomorphic to M by the mapping $y = x + m$ (x fixed, $m \in M$, $y \in [x]$). If $\|[x]\| = 0$ there is a sequence $\{y_n\}$ such that $y_n \in [x]$ and $\|y_n\| \to 0$. But then $y_n \to 0$, and so $0 \in [x]$, which means that $x \in M$. Thus $\|[x]\|$ is a norm on X/M.

Now suppose that X is a Banach space, that is, that X is complete as a metric space. To show that X/M is complete it suffices to show that every Cauchy sequence in X/M contains a convergent subsequence. For, if a subsequence of a Cauchy sequence has a limit, it is an easy matter to verify that the sequence itself has this limit. Let $\{U_n\}$ denote a Cauchy sequence in X/M. By induction it is possible to select a subsequence $\{U_{n(i)}\}$ such that $\|U_{n(i+1)} - U_{n(i)}\| < 1/2^i$, $i = 1, 2, \cdots$. For convenience we write $V_i = U_{n(i)}$. Now select $w_i \in V_{i+1} - V_i$ (this is a vector difference, not a set-theoretic difference) so that $\|w_i\| < 1/2^{i-1}$, $i = 1, 2, \cdots$. This is possible by the way in which the norm is defined in X/M. Select $v_1 \in V_1$, and, if $v_1, \cdots, v_i$, have been chosen, choose $v_{i+1} \in V_{i+1}$ so that

$w_i = v_{i+1} - v_i$. Then, if $m < n$, $v_n - v_m = (v_n - v_{n-1}) + (v_{n-1} - v_{n-2}) + \cdots + (v_{m+1} - v_m)$, $\|v_n - v_m\| \leqslant \|w_{n-1}\| + \|w_{n-2}\| + \cdots + \|w_m\| < \frac{1}{2^{m-1}} + \cdots + \frac{1}{2^{n-2}} < \frac{1}{2^{m-2}}$. Therefore $\{v_n\}$ is a Cauchy sequence in X, and $v = \lim\limits_{n\to\infty} v_n$ exists. Let $V = [v]$. Then

$$\|V_n - V\| = \|[v_n - v]\| \leqslant \|v_n - v\|,$$

so that $V_n \to V$. This finishes the proof that X/M is complete.

For instances of the use of the quotient-space concept, see Theorem 4.3–F and problems 5, 6 in § 4.6.

3.2 Inner-Product Spaces

Definition. A complex linear space X is called an inner-product space if there is defined on $X \times X$ a complex-valued function (x_1, x_2) (called the *inner product* of x_1 and x_2) with the following properties:

1. $(x_1 + x_2, x_3) = (x_1, x_3) + (x_2, x_3)$.
2. $(x_1, x_2) = \overline{(x_2, x_1)}$ (the bar denoting complex conjugate).
3. $(\alpha x_1, x_2) = \alpha(x_1, x_2)$.
4. (x, x) is $\geqslant 0$ [it must be real by (2)], and $(x, x) \neq 0$ if $x \neq 0$.

A *real* linear space X is called an inner-product space if there is defined on $X \times X$ a real-valued function (x_1, x_2) with the properties 1–4, except that (2) is to be written without the bar over (x_2, x_1).

Observe that, as a consequence of the properties listed, (x_1, x_2) has the further properties:

$$(x_1, x_2 + x_3) = (x_1, x_2) + (x_1, x_3),$$
$$(x_1, \alpha x_2) = \bar{\alpha}(x_1, x_2),$$

where the bar over α is to be omitted in the case of a real space.

There are fifteen theorems in this section; the last twelve of them refer to orthonormal systems. Of these latter theorems, the principal ones are 3.2–D (the generalized Bessel inequality), 3.2–G (the abstract counterpart of the classical Riesz-Fischer theorem), 3.2–H [one form of the projection theorem; see also problem 4, § 4.81, and § 4.82], 3.2–K (the Parseval identity), and 3.2–L (the Gram-Schmidt orthogonalization process).

The first important fact to be noted about inner-product spaces is that we can use the inner product to define a norm. We must first prove the following theorem.

Theorem 3.2–A. *If X is an inner-product space, then*

$$|(x_1, x_2)| \leqslant \sqrt{(x_1, x_1)}\,\sqrt{(x_2, x_2)}. \tag{3.2–1}$$

The equality holds in (3.2–1) *if and only if x_1 and x_2 are linearly dependent.*

PROOF. For any α and β we have $(\alpha x + \beta y, \alpha x + \beta y) = \alpha\bar{\alpha}(x, x) + \alpha\bar{\beta}(x, y) + \bar{\alpha}\beta(y, x) + \beta\bar{\beta}(y, y) \geqslant 0$. The idea of the proof is to reduce this to an inequality of the form $At^2 + 2Bt + C \geqslant 0$, with A, B, C real constants and t a real variable. We choose $\alpha = t$ (real) and define

$$\beta = \frac{(x, y)}{|(x, y)|} \qquad \text{if } (x, y) \neq 0,$$

$$\beta = 1, \text{ otherwise.}$$

Thus $\bar{\beta}(x, y) = |(x, y)|$ and $\beta\bar{\beta} = 1$. Therefore

$$t^2(x, x) + 2t|(x, y)| + (y, y) \geqslant 0 \tag{3.2–2}$$

for all real t. We conclude that

$$|(x, y)|^2 - (x, x)(y, y) \leqslant 0,$$

which is equivalent to (3.2–1). The equality $|(x, y)|^2 = (x, x)(y, y)$ can hold when $x \neq 0$ if and only if there is a real value of t for which the quadratic form in (3.2–2) takes on the value 0. But, because of the origin of the quadratic form, we see that this occurs if and only if $tx + \beta y = 0$ for some real t and some β such that $|\beta| = 1$. This relation implies and is implied by the statement that x and y are linearly dependent, provided that $x \neq 0$. When $x = 0$ the equality holds in (3.2–1), of course, and also x and y are linearly dependent. Thus the proof is complete.

The inequality (3.2–1) is known as *Schwarz's inequality.*

Theorem 3.2–B. *If X is an inner-product space, $\sqrt{(x, x)}$ has the properties of a norm.*

PROOF. We write $\|x\| = \sqrt{(x, x)}$. The only requisite property of $\|x\|$ that is not immediately apparent is the triangular inequality property. Now

$$\|x + y\|^2 = (x, x) + (x, y) + (y, x) + (y, y).$$

For complex A we write $\operatorname{Re} A = {}^1\!/_2(A + \bar{A})$ = real part of A. Since $|\operatorname{Re} A| \leqslant |A|$, we see that

$$\|x + y\|^2 \leqslant \|x\|^2 + 2|(x, y)| + \|y\|^2.$$

With (3.2–1) this gives

$$\|x + y\|^2 \leqslant \|x\|^2 + 2\|x\|\,\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2.$$

Consequently $\|x + y\| \leqslant \|x\| + \|y\|$, and the proof is complete.

While the norm is expressible in terms of the inner product, it is also true that the inner product is expressible in terms of the norm. If X is a *real* inner-product space, we have

$$(3.2\text{–}3) \qquad (x, y) = \tfrac{1}{4}[\|x + y\|^2 - \|x - y\|^2].$$

If X is a *complex* inner-product space, the real and imaginary parts of (x, y) are expressible in the forms

$$(3.2\text{–}4) \qquad \operatorname{Re}(x, y) = \tfrac{1}{4}[\|x + y\|^2 - \|x - y\|^2],$$

$$(3.2\text{–}5) \qquad \operatorname{Im}(x, y) = -\tfrac{1}{4}[\|ix + y\|^2 - \|ix - y\|^2].$$

The validity of these formulas is readily verified by direct simplification of the expressions on the right, using the properties of the inner product.

Throughout the rest of this section the discussion and theorems are concerned with inner-product spaces, and we shall not repeat this fact every time in stating theorems. It is furthermore assumed that an inner-product space is given the topology of a normed linear space, using the norm $\|x\| = \sqrt{(x, x)}$.

Theorem 3.2–C. *The inner product* (x_1, x_2) *is a continuous function on* $X \times X$.

PROOF. Let $x_1 - u_1 = v_1$, $x_2 - u_2 = v_2$, where x_1, x_2, u_1, u_2 are arbitrary. Then

$$\begin{aligned}(x_1, x_2) - (u_1, u_2) &= (u_1 + v_1, u_2 + v_2) - (u_1, u_2)\\ &= (u_1, v_2) + (v_1, u_2) + (v_1, v_2).\end{aligned}$$

Therefore, using (3.2–1) and the definitions of v_1, v_2, we have

$$\begin{aligned}|(x_1, x_2) - (u_1, u_2)| \leqslant \|u_1\|\,\|x_2 - u_2\|\\ + \|x_1 - u_1\|\,\|u_2\| + \|x_1 - u_1\|\,\|x_2 - u_2\|.\end{aligned}$$

The asserted continuity is now evident.

Examples of inner-product spaces. The space l^2 is an inner-product space, with inner product

$$(x, y) = \sum_{n=1}^{\infty} \xi_n \bar{\eta}_n,$$

where $x = \{\xi_n\}$, $y = \{\eta_n\}$.

The space $L^2(a, b)$ is an inner-product space, with inner product

$$(x, y) = \int_a^b x(t)\,\overline{y(t)}\,dt.$$

The two spaces just mentioned are complete. We can get many examples of *incomplete* inner-product spaces by selecting nonclosed linear manifolds in the aforementioned spaces. For instance, the subspace of $L^2(a, b)$ which is composed of elements corresponding to functions continuous on $[a, b]$ is an incomplete inner-product space.

The n-dimensional space $l^2(n)$ is an inner-product space, with

$$(x, y) = \sum_{k=1}^{n} \xi_k \bar{\eta}_k.$$

One of the most important notions in an inner-product space is that of orthogonality.

Definition. We say that x and y are *orthogonal* if $(x, y) = 0$. The statement "x and y are orthogonal" is often expressed symbolically in the form $x \perp y$.

Observe that $x \perp y$ is equivalent to $y \perp x$; also, $x \perp 0$ for every x.

Definition. A set S of vectors is called an *orthogonal set* if $x \perp y$ for every pair x, y for which $x \in S$, $y \in S$, and $x \neq y$. If in addition $\|x\| = 1$ for every $x \in S$, the set is called an *orthonormal set.*

The remainder of this section is devoted to a discussion of orthonormal sets. One of the main goals of the discussion is the determination of the nature of the closed linear manifold generated by a given orthonormal set: How is it related to the whole space? How may an element of the manifold be expressed in terms of the elements of the orthonormal set? We find here the inequality of Bessel and the Parseval identity. These are the general forms of the relations bearing these names which occur in the theory of Fourier series.

Theorem 3.2–D. *Let S be an orthonormal set. If $u_1, \cdots, u_n$ is any finite collection of distinct elements of S, and $x \in X$, then*

$$\sum_{i=1}^{n} |(x, u_i)|^2 \leqslant \|x\|^2. \tag{3.2–6}$$

The set of those $u \in S$ such that $(x, u) \neq 0$ (x any fixed element of X) is either finite or countably infinite. If $x, y \in X$,

$$\sum_{u \in S} |(x, u)\,\overline{(y, u)}| \leqslant \|x\|\,\|y\|, \tag{3.2–7}$$

it being understood that the sum on the left includes all $u \in S$ for which $(x, u)\overline{(y, u)} \neq 0$, and is, therefore, either a finite series or an absolutely convergent series with a countable infinity of terms.

PROOF. To prove (3.2–6) we write $\xi_i = (x, u_i)$. Then

$$0 \leqslant \left(x - \sum_1^n \xi_i u_i, x - \sum_1^n \xi_j u_j\right)$$

$$= (x, x) - \sum_1^n \bar{\xi}_j (x, u_j) - \sum_1^n \xi_i (u_i, x) + \sum_i \sum_j \xi_i \bar{\xi}_j (u_i, u_j).$$

But $(u_i, u_j) = \delta_{ij}$, since S is orthonormal. If we take into account the definition of ξ_i, we see that

$$0 \leqslant \|x\|^2 - \sum_j |\xi_j|^2 - \sum_i |\xi_i|^2 + \sum_i |\xi_i|^2 = \|x\|^2 - \sum_j |\xi_j|^2,$$

and this is equivalent to (3.2–6).

It follows from (3.2–6) that, if $x \in X$ and n is a positive integer, the number of elements u of S such that $|(x, u)| \geqslant 1/n$ cannot exceed $n^2\|x\|^2$. Since $|(x, u)| \geqslant 1/n$ for some n if $(x, u) \neq 0$, the set of those $u \in S$ such that $(x, u) \neq 0$ is a countable union of finite sets and is, therefore, either finite or countably infinite. The inequality (3.2–7) now follows from (3.2–6), for, if $u_1, \cdots, u_n$ is a finite collection of distinct elements of S, we have

$$\sum_{i=1}^n |(x, u_i)(\overline{y, u_i})| \leqslant \left(\sum_{i=1}^n |(x, u_i)|^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^n |(y, u_i)|^2\right)^{\frac{1}{2}} \leqslant \|x\| \, \|y\|,$$

by (3.2–6) and Cauchy's inequality.

Examples of orthonormal sets. In the space l^2 let $e_k (k = 1, 2, \cdots)$ be that element $x = \{\xi_n\}$ for which $\xi_n = 0$ if $n \neq k$, and $\xi_k = 1$. The set of all the e_k's, or any nonempty subset of this set, is an orthonormal set.

In the complex space $L^2(0, 2\pi)$ let u_n be the element corresponding to the function $(1/\sqrt{2\pi})e^{int} (n = 0, \pm 1, \pm 2, \cdots)$. Since

$$\frac{1}{2\pi} \int_0^{2\pi} e^{imt} \, e^{-int} \, dt = \begin{cases} 1 & \text{if } m = n \\ 0 & \text{if } m \neq n \end{cases}$$

we see that the various elements u_n form an orthonormal set.

If we deal with the *real* space $L^2(0, 2\pi)$, the elements determined by the functions

$$\frac{1}{\sqrt{2\pi}}, \quad \frac{1}{\sqrt{\pi}} \cos t, \quad \frac{1}{\sqrt{\pi}} \cos 2t, \quad \cdots$$

$$\frac{1}{\sqrt{\pi}} \sin t, \quad \frac{1}{\sqrt{\pi}} \sin 2t, \quad \cdots$$

form an orthonormal set.

Other examples of orthonormal sets are mentioned in § 3.22.

Theorem 3.2–E. *Suppose X is separable and that S is an orthonormal set in X. Then S is either a finite or a countably infinite set.*

PROOF. If $x \perp y$ and $\|x\| = \|y\| = 1$, we have

$$(x - y, x - y) = (x, x) - (x, y) - (y, x) + (y, y) = 2,$$

or $\|x - y\| = \sqrt{2}$. Thus, in an orthonormal set, each pair of distinct elements are a distance $\sqrt{2}$ apart. Now let $\{y_n\}$ be a denumerable set everywhere dense in X. Then to each $x \in S$ corresponds some n, depending on x, such that $\|x - y_n\| < \sqrt{2}/3$. If x and u are distinct members of S and if $\|u - y_m\| < \sqrt{2}/3$, then

$$\sqrt{2} = \|x - u\| \leqslant \|x - y_n\| + \|y_n - y_m\| + \|y_m - u\|$$
$$< \frac{2\sqrt{2}}{3} + \|y_n - y_m\|,$$

or

$$\frac{\sqrt{2}}{3} < \|y_n - y_m\|,$$

so that $n \neq m$. There is thus a one-to-one correspondence between the elements of the orthonormal set S and the elements of a subset of the countable set $\{y_n\}$. This proves the theorem.

Theorem 3.2–F. *Let $u_1, \cdots, u_n$ be a finite orthonormal set in the space X, and let M be the subspace of X generated by $u_1, \cdots, u_n$. Then $u_1, \cdots, u_n$ is a basis for M, and the coefficients in a representation $x = \xi_1 u_1 + \cdots + \xi_n u_n$ of an element of M are related to x by the formulas $\xi_i = (x, u_i)$.*

PROOF. If $x = \xi_1 u_1 + \cdots + \xi_n u_n$, then

$$(x, u_i) = \xi_1(u_1, u_i) + \cdots + \xi_n(u_n, u_i) = \xi_i,$$

by the orthonormality relations. If $x = 0$ it follows that $\xi_1 = \cdots = \xi_n = 0$. Hence we see that the u_i's form a linearly independent set and, hence, form a basis for the subspace which they generate.

Theorem 3.2–G. *Suppose X is complete, and let $\{u_n\}$ be a countably infinite orthonormal set in X. Then a series of the form $\sum_1^\infty \xi_n u_n$ is convergent if and only if $\sum_1^\infty |\xi_n|^2 < \infty$, and in that case we have the relations*

$$\xi_n = (x, u_n), \qquad x = \sum_1^\infty \xi_n u_n,$$

between the coefficients ξ_n and the element x defined by the series.

PROOF. Let $s_n = \xi_1 u_1 + \cdots + \xi_n u_n$. Then, if $m < n$, the orthonormality relations and the relation between the norm and inner product lead to the formula

$$\|s_n - s_m\|^2 = \left\| \sum_{i=m+1}^{n} \xi_i u_i \right\|^2 = \sum_{i=m+1}^{n} |\xi_i|^2.$$

Since X is complete, it is now clear that the sequence $\{s_n\}$ is convergent in X if and only if $\sum_1^\infty |\xi_i|^2 < \infty$. If this latter condition is satisfied, and if $x = \lim_{n\to\infty} s_n$, we prove that $\xi_i = (x, u_i)$ as follows: By Theorem 3.2–F we know that $\xi_i = (s_n, u_i)$ if $1 \leqslant i \leqslant n$. But $s_n \to x$ and hence $(s_n, u_i) \to (x, u_i)$ when $n \to \infty$, by the continuity of the inner product. Therefore $\xi_i = (x, u_i)$. This completes the proof.

The Riesz-Fischer theorem in the theory of Fourier series is a concrete instance of Theorem 3.2–G. This theorem asserts that, if $a_0, a_1, a_2, \cdots$ and $b_1, b_2, \cdots$ are sequences of real constants such that

$$\frac{a_0^2}{2} + \sum_1^\infty (a_n^2 + b_n^2) < \infty,$$

there exists a function x of class $\mathscr{L}^2(0, 2\pi)$ having a_n and b_n as its Fourier coefficients, i.e.,

$$a_n = \frac{1}{\pi}\int_0^{2\pi} x(t)\cos nt\, dt, \qquad b_n = \frac{1}{\pi}\int_0^{2\pi} x(t)\sin nt\, dt.$$

We take X to be $L^2(0, 2\pi)$ and for $u_1, u_2, u_3, u_4, u_5, \cdots$ the elements of X corresponding to

$$\frac{1}{\sqrt{2\pi}}, \quad \frac{\cos t}{\sqrt{\pi}}, \quad \frac{\sin t}{\sqrt{\pi}}, \quad \frac{\cos 2t}{\sqrt{\pi}}, \quad \frac{\sin 2t}{\sqrt{\pi}}, \cdots.$$

The coefficients $\xi_1, \xi_2, \cdots$ in Theorem 3.2–G are related to the standard Fourier coefficients by the formulas

$$\xi_1 = a_0\sqrt{\pi/2}, \qquad \xi_2 = a_1\sqrt{\pi}, \qquad \xi_3 = b_1\sqrt{\pi}, \cdots.$$

If x is orthogonal to each element of a set S, we say that x is orthogonal to S, and write $x \perp S$. Since the inner product is continuous, it follows that if $x \perp y_n$ and $y_n \to y$, then also $x \perp y$. Also, if $x \perp y$ and $x \perp z$, then $x \perp (y + z)$ and $x \perp (\alpha y)$ for every scalar α. Hence, if $x \perp S$, then x is also orthogonal to the linear manifold generated by S, and to the closure of this linear manifold.

Theorem 3.2–H. *Let S be an orthonormal set in the space X. If S is* *infinite, assume further that X is complete. For each $x \in X$ there is an element of X unambiguously defined by*

$$x_S = \sum_{u \in S} (x, u)u.$$

Let M be the closed manifold generated by S. Then $x \in M$ if and only if $x = x_S$. In any case, $x_S \in M$ and $(x - x_S) \perp M$.

PROOF. If $x \in X$, we know by Theorem 3.2–D that $(x, u) \neq 0$ for at most a countably infinite number of elements $u \in S$. If we index the u's for which $(x, u) \neq 0$, in some arbitrary order, say $u_1, u_2, \cdots$, the series $\sum_n (x, u_n)u_n$, if infinite, is convergent (by Theorems 3.2–D and 3.2–G). Moreover, the series remains convergent, no matter how its terms are rearranged, as may be seen by the first part of the proof of Theorem 3.2–G, using the fact that the series $\sum_n |(x, u_n)|^2$ is absolutely convergent. We may then show that the series $\sum_n (x, u_n)u_n$ converges to the same element, no matter how the terms are rearranged. For, if $\{v_n\}$ is a rearrangement of $\{u_n\}$, and

$$x_1 = \sum_n (x, u_n)u_n, \qquad x_2 = \sum_n (x, v_n)v_n,$$

we have $(x_1, u_n) = (x, u_n)$ and $(x_2, v_n) = (x, v_n)$, by Theorem 3.2–G. Thus, if $u_n = v_{m(n)}$, we have $(x_1 - x_2, u_n) = (x_1, u_n) - (x_2, u_n) = (x, u_n) - (x_2, v_{m(n)}) = (x, u_n) - (x, v_{m(n)}) = (x, u_n) - (x, u_n) = 0$, and from this it follows, on setting $\alpha_n = (x, u_n)$, $\beta_n = (x, v_n)$, that

$$\begin{aligned} \|x_1 - x_2\|^2 &= \left(x_1 - x_2, \sum_n \alpha_n u_n - \sum_n \beta_n v_n\right) \\ &= \sum_n \bar{\alpha}_n (x_1 - x_2, u_n) - \sum_n \bar{\beta}_n (x_1 - x_2, v_n) = 0, \end{aligned}$$

whence $x_1 = x_2$. These considerations show us that the notation $\sum_{u \in S} (x, u)u$ has an unambiguous meaning and defines an element which we denote by x_S. Evidently $x_S \in M$. To prove $(x - x_S) \perp M$ it suffices to prove $(x - x_S) \perp S$. Let v be an arbitrary element of S. Then

$$\begin{aligned} (x - x_S, v) &= (x, v) - \sum_{u \in S} (x, u)(u, v) \\ &= (x, v) - (x, v) = 0, \end{aligned}$$

and hence $(x - x_S) \perp S$.

Since $x_S \in M$, it remains only to prove that $x_S = x$ if $x \in M$. Now, if $x \in M$, then $x - x_S \in M$, and since $(x - x_S) \perp M$, we have

$$\|x - x_S\|^2 = (x - x_S, x - x_S) = 0,$$

whence $x = x_S$. It is clear from a scrutiny of the proof that we need not assume the completeness of X if S is a finite set, for in that case no convergence questions arise, and the linear manifold generated by S, being finite dimensional, is closed.

Definition. An orthonormal set S in the space X is called *complete* if there exists no orthonormal set of which S is a proper subset. In other words, S is called complete if it is maximal with respect to the property of being orthonormal.

Theorem 3.2–I. *Every inner-product space X having a nonzero element contains a complete orthonormal set. Moreover, if S is any orthonormal set in X, there is a complete orthonormal set containing S as a subset.*

PROOF. Let S be an orthonormal set in X. Such sets certainly exist; for instance, if $x \neq 0$, the set consisting merely of $x/\|x\|$ is orthonormal. Let P be the class of all orthonormal sets having S as a subset. It is easily verified that P satisfies the conditions of Zorn's lemma (§ 1.7). Thus P contains a maximal member.

Theorem 3.2–J. *Let S be an orthonormal set in X, and let M be the closed linear manifold generated by S. If $M = X$, it follows that S is complete. If the space X is complete and the orthonormal set S is complete (i.e., maximal), then $M = X$.*

PROOF. If S is not complete, there exists some $x \neq 0$ such that $x \perp S$, and hence also $x \perp M$. Now, if $M = X$, we have $x \perp x$, which implies $x = 0$. Thus, if S is not complete, we must conclude that $M \neq X$. This proves the first part of the theorem. Let us now assume that X is complete. Then, if $M \neq X$, suppose $x \in X - M$ and construct x_S as in Theorem 3.2–H. Then let

$$y = \frac{x - x_S}{\|x - x_S\|}.$$

Then the set consisting of S and y is orthonormal, and $y \notin S$, so that S cannot be complete. This completes the proof.

The assumption that X is complete in the second part of the theorem is essential, for it has been shown by an example that, if X is incomplete, it can happen that there is no orthonormal set in X such that the closed linear manifold generated by the set is all of X (see Dixmier, **2**).

Example of a complete orthonormal set. We have already mentioned the orthonormal set composed of $e_1, e_2, e_3, \cdots$ in the space l^2. This is a complete orthonormal set, as we now prove. If $x = \{\xi_n\}$ let $x_n = \xi_1 e_1 + \cdots + \xi_n e_n$. Then

$$\|x - x_n\|^2 = \sum_{k=n+1}^{\infty} |\xi_k|^2,$$

and so $\|x - x_n\| \to 0$. It follows that the closed linear manifold generated by the elements $e_1, e_2, \cdots$ is all of l^2. Hence, by Theorem 3.2–J the orthonormal set is complete. In this example we observe that $\xi_n = (x, e_n)$ and hence that

$$\|x\|^2 = \sum_{n=1}^{\infty} |(x, e_n)|^2.$$

In the next theorem we see that this formula is a special case of a relation which is characteristic of complete orthonormal sets, provided we deal with complete spaces.

Theorem 3.2–K. *Let S be an orthonormal set in X. Suppose that*

$$\|x\|^2 = \sum_{u \in S} |(x, u)|^2 \tag{3.2–8}$$

for every $x \in X$. Then S is complete. On the other hand, if we assume that X is a complete space and that S is a complete orthonormal set, it follows that (3.2–8) *holds true for every $x \in X$.*

PROOF. If S is not complete, there exists some $x \neq 0$ such that $x \perp S$. But (3.2–8) then gives $\|x\|^2 = 0$. This proves the first assertion. Now suppose that X is complete (as a metric space) and that S is complete (as an orthonormal set). For each $x \in X$ there is some finite or countable set $\{u_n\}$ in S such that

$$x = \sum_{u \in S} (x, u)u = \sum_n \xi_n u_n, \qquad \xi_n = (x, u_n).$$

Here we use Theorems 3.2–H and 3.2–J. Now

$$\|x\|^2 = \left(\sum_n \xi_n u_n, \sum_m \xi_m u_m\right) = \sum_n \sum_m \xi_n \bar{\xi}_m (u_n, u_m) = \sum_n |\xi_n|^2,$$

by the orthonormality relations. The equality just obtained is equivalent to (3.2–8).

The equality (3.2–8) is called *Parseval's formula.* The inequality

$$\sum_{u \in S} |(x, u)|^2 \leqslant \|x\|^2, \tag{3.2–9}$$

which holds for *all* orthonormal systems, complete or not (see Theorem 3.2–D), is called *Bessel's inequality.*

The same method which was used to derive (3.2–8) may be used to prove Parseval's formula in the more general form

$$(x, y) = \sum_{u \in S} (x, u)\overline{(y, u)}, \tag{3.2–10}$$

where S is a complete orthonormal set, and x, y are arbitrary elements of X. It is assumed that X is a complete space.

There is an important constructive process for obtaining an orthonormal set from a finite or countable linearly independent set.

Theorem 3.2–L. *Suppose $\{x_n\}$ is a finite or countable linearly independent set. Then, there exists an orthonormal set having the same cardinal number and generating the same linear manifold as the given set.*

PROOF. Certainly $x_1 \neq 0$. We define $y_1, y_2, \cdots$ and $u_1, u_2, \cdots$ recursively, as follows:

$$\begin{aligned}
&y_1 = x_1 && u_1 = \frac{y_1}{\|y_1\|},\\
&y_2 = x_2 - (x_2, u_1)u_1 && u_2 = \frac{y_2}{\|y_2\|},\\
&\cdots && \cdots\\
&y_{n+1} = x_{n+1} - \sum_{k=1}^{n} (x_{n+1}, u_k)u_k, && u_{n+1} = \frac{y_{n+1}}{\|y_{n+1}\|},\\
&\cdots && \cdots
\end{aligned}$$

The process terminates if the set of x's is finite. Otherwise it continues indefinitely. It is clear by induction that u_n is a linear combination of $x_1, \cdots, x_n$, and vice versa. Thus, the linear manifold generated by the u's is the same as that generated by the x's. Observe that $y_n \neq 0$, because the set $x_1, \cdots, x_n$ is linearly independent. By direct calculation we see that $(y_{n+1}, u_i) = 0$ if $i = 1, \cdots, n$. From this it follows that the u's form an orthonormal set.

This process of constructing the u's from the x's is called the Gram-Schmidt orthogonalization process.

For a finite-dimensional space we obtain the following result:

Theorem 3.2–M. *If X is a space of finite dimension $n \geqslant 1$, it has a basis which is an orthonormal set. Moreover, every complete orthonormal set in X is a basis.*

PROOF. If $x_1, \cdots, x_n$ is any basis of X, the Gram-Schmidt process gives us an orthonormal basis $u_1, \cdots, u_n$. An orthonormal set is linearly independent, as we saw in the proof of Theorem 3.2–F. It follows from Theorem 3.2–J that a complete orthonormal set in X is necessarily a basis.

Theorem 3.2–N. *Suppose X is infinite dimensional and separable (though not necessarily complete). Then, there exists in X a countable complete orthonormal set S such that the closed linear manifold generated by S is X.*

PROOF. Since X is separable, there exists a countable set $\{x_n\}$ which is everywhere dense in X. Let y_1 be the first nonzero element in the sequence $\{x_n\}$, y_2 the first x_n which is not in the linear manifold generated by y_1, and y_{k+1} the first x_n which is not in the linear manifold generated by $y_1, \cdots, y_k$. It is clear that the x's and the y's generate the same linear manifold and hence the same closed linear manifold. This manifold is the whole space X, because the set of x's is dense in X. Applying the Gram-Schmidt process to the y's, we obtain an orthonormal set which generates the closed linear manifold X and which must therefore be complete, by the first part of Theorem 3.2–J. The orthonormal set cannot be finite, since X is not finite dimensional.

The assumption in Theorem 3.2–N that X is separable is essential, even if we omit the word "countable" in the conclusion. This can be shown by an example; see the remarks following the proof of Theorem 3.2–J.

Theorem 3.2–O. *Let X be an inner-product space, and let S_1 and S_2 be two complete orthonormal sets in X. Then S_1 and S_2 have the same cardinal number (i.e., there is a one-to-one correspondence between the elements of S_1 and S_2).*

PROOF. Suppose that one of the sets S_1, S_2, say S_1, is finite, and let M be the closed linear manifold determined by S_1. Then $M = X$, for otherwise the set S_1 would not be complete, as we can see readily with the aid of Theorem 3.2–H. Hence X is finite dimensional. But then S_1 and S_2 are both bases for X, by Theorem 3.2–M, and the sets S_1, S_2 must both have the same finite number of elements. For the remainder of the proof we therefore assume that S_1 and S_2 are both infinite sets. For each $x \in S_1$ let $S_2(x)$ be the set of those elements $y \in S_2$ such that $(x, y) \neq 0$. We know by Theorem 3.2–D that $S_2(x)$ is an at most countably infinite set. If $y \in S_2$, then $\|y\| = 1$, and it follows from the completeness of S_1 that there is some $x \in S_1$ such that $y \in S_2(x)$ (since otherwise S_1 together with y would be an orthonormal set). Hence we see that

$$S_2 = \bigcup_{x \in S_1} S_2(x). \tag{3.2–11}$$

Now let A_1, A_2 be the cardinal numbers of the sets S_1, S_2 respectively, and let A_0 be the cardinal number of the set of positive integers. The relation (3.2–11) shows that $A_2 \leqslant A_0 A_1$. But it is known that, if A is any infinite cardinal number, then $A_0 A = A$ (see, for example, Sierpinski, **1**, § 103). Thus, we see that $A_2 \leqslant A_1$. So far we have used only the completeness of S_1. If S_2 is also complete, we can exchange the roles of S_1 and S_2 and so conclude $A_1 \leqslant A_2$. But then $A_1 = A_2$, and the proof is finished.

PROBLEM

If X is a complex complete inner-product space, there exists an operator C of the following sort: The domain of C is X, the range of C is in X, and C has the properties

$$C(x + y) = Cx + Cy, \qquad C(\alpha x) = \bar{\alpha} Cx,$$
$$C^2 x = x, \qquad \|Cx\| = \|x\|.$$

Such an operator may be called a *conjugation* of X, and the notation $\bar{x} = Cx$ is appropriate. The further property $(Cx, Cy) = \overline{(x, y)}$ may be deduced, using (3.2–4) and (3.2–5). Furthermore, the range of C is *all* of X, and C has the inverse $C^{-1} = C$. To obtain such an operator, let S be a complete orthonormal set in X, and if $x = \sum_{u \in S} (x, u)u$ is the representation of x, let $Cx = \sum_{u \in S} \overline{(x, u)}u$.

3.21 Hilbert Spaces

Definition. An inner-product space which is infinite dimensional and complete is called a *Hilbert space* (real or complex, according to whether the scalar field is real or complex). If X is a finite-dimensional inner-product space, we shall call it a *Euclidean space.*

In the early literature on Hilbert spaces, it was usual to assume separability as part of the definition of a Hilbert space, and separability was used in proving some of the most important theorems. But it was subsequently discovered that these theorems could be proved without the hypothesis of separability, and the current practice is to invoke this hypothesis when it is needed, but not otherwise.

Some writers have used the adjective *Euclidean* for any complete inner-product space, or even for incomplete inner-product space. Also, *complex* inner-product spaces of finite dimension are frequently called *unitary* spaces.

An incomplete inner-product space X, being a normed linear space, can be completed in the manner described in § 3.13. Moreover, if Y is the completion of X, we can define an inner product on $Y \times Y$ by

$$(y, z) = \lim_{n\to\infty} (x_n, u_n),$$

where y, z are equivalence classes of Cauchy sequences from X containing $\{x_n\}$ and $\{u_n\}$, respectively. The existence of the limit defining (y, z) may be seen with the aid of the formulas (3.2–3)–(3.2–5). We leave it for the reader to take care of the details of showing that (y, z) has all the requisite properties of an inner product. We see then that any incomplete inner-product space X can be extended to form a Hilbert space in which X is everywhere dense. Because of this fact, an incomplete inner-product space may be descriptively called a *pre-Hilbert space.* This terminology has been used by a number of writers.

We now give some theorems about concrete representations of Euclidean and Hilbert spaces.

Theorem 3.21–A. *Every Euclidean space X of finite dimension $n(n \geqslant 1)$ is congruent to $l^2(n)$.*

PROOF. For the definition of congruence see § 3.1. We can choose for X an orthonormal basis $u_1, \cdots, u_n$. Then, if $x = \xi_1 u_1 + \cdots + \xi_n u_n$, the correspondence $x \leftrightarrow (\xi_1, \cdots, \xi_n)$ establishes the congruence of X and $l^2(n)$. The isometric character of the correspondence follows by direct calculation, for, if $y = \eta_1 u_1 + \cdots + \eta_n u_n$, we have

$$(x, y) = \xi_1\bar{\eta}_1 + \cdots + \xi_n\bar{\eta}_n.$$

Hence

$$\|x\| = (|\xi_1|^2 + \cdots + |\xi_n|^2)^{\frac{1}{2}},$$

and the right member of this equation is the norm of $(\xi_1, \cdots, \xi_n)$ in $l^2(n)$.

It is interesting to observe that, if two inner-product spaces X and Y are congruent and if T is the operator which maps X isomorphically and isometrically onto Y, we not only have $\|Tx\| = \|x\|$, but also $(Tx_1, Tx_2) = (x_1, x_2)$, for every pair x_1, x_2 in X. That this is so follows directly from formulas (3.2–3)–(3.2–5).

Theorem 3.21–B. *Let X be a separable Hilbert space. Then X is congruent to l^2.*

PROOF. We know (Theorem 3.2–N) that X contains a countable complete orthonormal set $\{u_n\}$. If $x \in X$, let $\xi_n = (x, u_n)$. Then the sequence $\{\xi_n\}$ belongs to l^2 and

$$\|x\|^2 = \sum_{n=1}^{\infty} |\xi_n|^2,$$

by Parseval's formula (3.2–8). Moreover, every $\{\xi_n\} \in l^2$ arises from some $x \in X$ in this way, by Theorem 3.2–G. The correspondence $x \leftrightarrow \{\xi_n\}$ clearly establishes an isometric and isomorphic correspondence between the elements of X and l^2, so that these spaces are congruent.

In order to get a result like that of Theorem 3.21–B for nonseparable Hilbert spaces we must first construct a space somewhat like l^2, but of such a character that it need not be separable.

Definition of the Space $l^2[Q]$. Let Q be any nonempty set of elements. Let $l^2[Q]$ be the class of all complex-valued functions x defined on Q such that the set of $q \in Q$ for which $x(q) \neq 0$ is either finite or countable and, moreover,

$$\sum_{q \in Q} |x(q)|^2 < \infty.$$

This class becomes a complex inner-product space if we define $x + y$ and αx as usual with functions and define the inner product by

$$(x, y) = \sum_{q \in Q} x(q)\overline{y(q)}.$$

Moreover, the space is complete. This is clear if Q is a finite set with n elements or if Q is countable, for in these cases $l^2[Q]$ is congruent to $l^2(n)$ and to l^2 respectively (in fact $l^2[Q]$ is $l^2(n)$ if Q is the set of integers $1, \cdots, n$ and $l^2[Q]$ is l^2 if Q is the set of *all* positive integers). If Q is uncountable, the fact that $l^2[Q]$ is complete follows readily from the fact that l^2 is complete (see § 3.13).

We have defined the *complex* space $l^2[Q]$; it is clear that, if we require all the functions x to be real-valued, we get a real inner-product space.

Now, for each $p \in Q$ let x_p be the function defined by $x_p(q) = 0$ if $p \neq q$, $x_p(p) = 1$. The set of all the x_p's is an orthonormal set; clearly its cardinal number is the same as that of the set Q. Hence, by Theorem 3.2–E, $l^2[Q]$ is not separable if Q is uncountable. The orthonormal set formed by the x_p's is complete. For, if $x \in l^2[Q]$, the definition of the inner product shows that $(x, x_p) = x(p)$, and thus, by the definition of the norm,

$$\|x\|^2 = \sum_{p \in Q} |(x, x_p)|^2.$$

Therefore the orthonormal set is complete, by Theorem 3.2–K.

This example shows that there exist Hilbert spaces having complete orthonormal sets whose cardinal number is any specified infinite cardinal.

Theorem 3.21–C. *Let X be a Hilbert space for which the cardinal number of any one (and hence of every) complete orthonormal set is A. Let Q be any set the cardinal number of whose elements is A. Then X is congruent to $l^2[Q]$.*

The proof is just like that of Theorem 3.21–B.

3.22 The Completeness of Certain Orthonormal Sets

Example 1. It was mentioned in § 3.2 that the functions $(1/\sqrt{2\pi})e^{int}$ $n = 0, \pm 1, \pm 2, \cdots$ determine an orthonormal set in the complex space $L^2(0, 2\pi)$. It is a fact of paramount importance in the theory of Fourier series that this orthonormal set is complete, i.e., that, if

$$\int_0^{2\pi} x(t)e^{int}\,dt = 0$$

for all integers n, then $x(t) \overset{0}{=} 0$. For a proof see Zygmund, **1**, § 1.5. By Theorem 3.2–K the completeness also finds its expression in the Parseval relation

$$\frac{1}{2\pi}\int_0^{2\pi} |x(t)|^2\,dt = \sum_{n=-\infty}^{\infty} |c_n|^2 \tag{3.22–1}$$

where

$$c_n = \frac{1}{2\pi}\int_0^{2\pi} x(t)e^{-int}\,dt.$$

The real functions

$$(1/\sqrt{2\pi}), \quad (1/\sqrt{\pi})\cos t, \quad (1/\sqrt{\pi})\cos 2t,$$
$$(1/\sqrt{\pi})\sin t, \quad (1/\sqrt{\pi})\sin 2t, \quad \cdots$$

also form a complete orthonormal system. The Parseval relation here (for real functions), in terms of the usual Fourier coefficients, is

$$\frac{1}{\pi}\int_0^{2\pi} \{x(t)\}^2\,dt = \frac{a_0^2}{2} + \sum_1^{\infty} (a_n^2 + b_n^2). \tag{3.22–2}$$

Example 2. In the space $L^2(-\infty, \infty)$ one example of a complete orthonormal system is furnished by the elements corresponding to the Hermite functions, which are defined in terms of the Hermite polynomials. These polynomials are

$$H_n(t) = (-1)^n e^{t^2} \frac{d^n}{dt^n} e^{-t^2}, \qquad n = 0, 1, 2, \cdots,$$

and the Hermite function ψ_n is defined as

$$\psi_n(t) = (2^n n!\,\sqrt{\pi})^{-\frac{1}{2}} H_n(t) e^{-t^2/2}.$$

The fact that the ψ_n's are pairwise orthogonal is a consequence of the fact that

$$\psi_n''(t) - t^2\psi_n(t) = -(2n + 1)\psi_n(t).$$

For proofs that the ψ_n's form a complete orthonormal set see Wiener, **1**, page 64, and Achieser and Glasmann, **1**, pages 27–28.

Example 3. Another interesting example is that of the Laguerre functions

$$\phi_n(t) = \frac{1}{n!} e^{-t/2} L_n(t), \qquad n = 0, 1, 2, \cdots$$

where

$$L_n(t) = e^t \frac{d^n}{dt^n} (t^n e^{-t})$$

is the Laguerre polynomial of degree n. The functions $\{\phi_n\}$ determine a complete orthonormal system in the space $L^2(0, \infty)$. See Courant and Hilbert, **1**, volume I, page 81; this proof shows that the closed linear manifold determined by the functions $\{\phi_n\}$ contains all the elements of $L^2(0, \infty)$ corresponding to piecewise continuous functions in $\mathscr{L}^2(0, \infty)$. Since elements of $L^2(0, \infty)$ corresponding to functions of this latter type are dense in $L^2(0, \infty)$ it follows by Theorem 3.2–J that the Laguerre functions determine a complete orthonormal system.

Example 4. The Legendre polynomial of degree n may be defined as

$$P_n(t) = \frac{1}{2^n n!} \frac{d^n}{dt^n} (t^2 - 1)^n.$$

The functions

$$\{\sqrt{n + \tfrac{1}{2}}\, P_n(t)\} \qquad n = 0, 1, 2, \cdots$$

determine a complete orthonormal set in the space $L^2(-1, 1)$. The argument in Churchill, **1**, pages 185–186, can be modified so as to prove this assertion.

We shall see later on (see Theorem 6.4–D and the latter part of § 5.5) that certain types of integral equations lead to the determination of complete orthonormal sets. This situation occurs, in particular, when certain types of boundary value problems for differential equations are recast as problems of integral equations (see § 6.41).

3.3 Topological Linear Spaces

The general definition of a topological linear space was given in § 3.0. In this section we give a rather brief discussion of such spaces, with emphasis upon the way in which the topology is determined by systems of neighborhoods of 0. Throughout the section X denotes a topological linear space.

The principal theorems are 3.3–E, 3.3–F, and 3.3–G. The first two of these theorems are concerned with fundamental aspects of systems of neighborhoods of 0 in a topological linear space; the third theorem brings out the important fact that in a topological linear space the weak T_1-separation axiom implies that the space is a Hausdorff space (and even somewhat more than this).

One important preliminary comment is this: The function f defined by $f(x) = \alpha x + x_0$, where α and x_0 are fixed, is continuous. Moreover, if $\alpha \neq 0$, then f is a homeomorphism of X onto all of X. It therefore maps open sets onto open sets. In particular, if S is a neighborhood of 0, $f(S)$ is a neighborhood of x_0.

If S, S_1, S_2 are sets in X, B is a set of scalars, x_0 is a fixed vector and α is a fixed scalar, we write

$$\begin{aligned} S_1 + S_2 &= \{x_1 + x_2 : x_1 \in S_1,\ x_2 \in S_2\}, \\ x_0 + S &= \{x_0 + x : x \in S\}, \\ BS &= \{\beta x : \beta \in B,\ x \in S\}, \\ \alpha S &= \{\alpha x : x \in S\}. \end{aligned}$$

Since $S_1 - S_2$ is regularly used for set-theoretic difference, it is necessary to distinguish carefully between $S_1 - S_2$ and $S_1 + (-1)\cdot S_2$. We *do* denote $(-1)\cdot S$ by $-S$, however, when this symbol stands alone.

Definition. A set S in X is called *symmetric* if $x \in S$ implies $-x \in S$, i.e., if $-S \subset S$. This is equivalent to $-S = S$.

Every neighborhood U of 0 contains a symmetric neighborhood of 0, e.g., $U \cap (-U)$.

Definition. A set S in X is called *balanced* (in the Bourbaki books the term is *équilibré*) if $ES \subset S$, where E is the set of all scalars α such that

$|\alpha| \leqslant 1$. If T is any nonempty set in X, the set ET is balanced. It is called the *balanced hull* of T. A balanced set coincides with its balanced hull.

The intersection of balanced sets is balanced. The balanced hull of S is the intersection of all balanced sets which contain S.

Observe that a balanced set is symmetric and contains 0.

The three following theorems are stated formally for purposes of later reference.

Theorem 3.3–A. *Let $\mathscr{U}$ be a base at 0 in X. Then, the family of sets $x_0 + U$, U varying over $\mathscr{U}$, is a base at x_0.*

The simple proof is left to the reader.

Theorem 3.3–B. *If $\mathscr{U}$ is a base at 0 and $U \in \mathscr{U}$, there exists $V \in \mathscr{U}$ such that $V + V \subset U$.*

PROOF. Use the facts that $0 + 0 = 0$ and that addition is continuous. We omit the details.

Theorem 3.3–C. *If S is a neighborhood of 0, the balanced hull of S is open.*

The simple proof is left to the reader. It may be observed that, if S is open but does not contain 0, the balanced hull of S may not be open. However, if the point 0 is deleted from the balanced hull, the set which remains is open.

The following theorem shows that there exists a family of balanced neighborhoods of 0 constituting a base at 0.

Theorem 3.3–D. *If $\mathscr{U}$ is a base at 0 and $\mathscr{V}$ is the family of balanced hulls of the members of $\mathscr{U}$, then $\mathscr{V}$ is a base at 0.*

PROOF. Suppose $U \in \mathscr{U}$. From $0 \cdot 0 = 0$ and the continuity of products it is possible to obtain a neighborhood of 0, say W_1, whose balanced hull W_2 is contained in U. Since W_2 is a neighborhood of 0 (by Theorem 3.3–C), there exists $U_1 \in \mathscr{U}$ such that $U_1 \subset W_2$. If V is the balanced hull of U_1, we have $V \subset W_2$, and hence $V \subset U$. This completes the proof.

Definition. A set S in X is called *absorbing* if to each $x \in X$ corresponds some $\epsilon > 0$ such that $\alpha x \in S$ if $0 < |\alpha| \leqslant \epsilon$. An equivalent formulation is: S is absorbing if to each $x \in X$ there corresponds some $r > 0$ such that $x \in \alpha S$ if $|\alpha| \geqslant r$.

It is clear at once from the continuity of products and the fact that $0 \cdot x = 0$ for every x that each neighborhood of 0 is absorbing.

The concepts of balanced sets and absorbing sets do not require a topology for their definition; they make sense in any linear space. As we shall presently see, a suitable family of balanced and absorbing sets in a linear space can be used to define a topology for the space.

A balanced set S is absorbing if and only if to each $x \in X$ there corresponds some $\alpha \neq 0$ such that $\alpha x \in S$.

Theorem 3.3–E. *If X is a topological linear space, there exists a family $\mathscr{U}$ of open sets forming a base at 0 and having the properties:*

1. *Each member of the family is balanced and absorbing.*
2. *If $U \in \mathscr{U}$ and $\alpha \neq 0$, then $\alpha U \in \mathscr{U}$.*
3. *If $U \in \mathscr{U}$, there exists $V \in \mathscr{U}$ such that $V + V \subset U$.*

PROOF. Let $\mathscr{U}$ be the family of all balanced neighborhoods of 0, which is the same as the family of balanced hulls of *all* neighborhoods of 0. We leave the verification of properties 1–3 to the reader.

In a normed linear space X the sets $\{x : \|x\| < r\}$, r varying over all positive numbers, form a family with the properties specified for $\mathscr{U}$ in Theorem 3.3–E. In proving theorems about topological linear spaces, a base with these properties is frequently used to pattern arguments after proofs using norms.

The next theorem shows how a topology may be constructed for a linear space.

Theorem 3.3–F. *Let X be a linear space. Let $\mathscr{U}$ be a nonempty family of nonempty subsets of X with the following properties:*

1. *Each member of $\mathscr{U}$ is balanced and absorbing.*
2. *If $U \in \mathscr{U}$, there exists $V \in \mathscr{U}$ such that $V + V \subset U$.*
3. *If U_1 and U_2 are in $\mathscr{U}$, there exists $U_3 \in \mathscr{U}$ such that $U_3 \subset U_1 \cap U_2$.*
4. *If $U \in \mathscr{U}$ and $x \in U$, there exists $V \in \mathscr{U}$ such that $x + V \subset U$.*

Then there is a unique topology for X such that X is a topological linear space with $\mathscr{U}$ as a base at 0.

PROOF. A nonempty balanced set contains 0. Hence $x \in x + U$ for each $U \in \mathscr{U}$. We shall define a set S to be open if to each $x \in S$ there corresponds some $U \in \mathscr{U}$ such that $x + U \subset S$. The elements of $\mathscr{U}$ are open, by (4), and (3) shows that the intersection of two open sets is open. The class of open sets, as defined, obviously has the remaining properties required in order for it to be a topology for X. The set $\mathscr{U}$ is evidently a base at 0.

The fact that addition is continuous follows at once from (2). Before proving continuity of multiplication we observe the following:

5. *If $U \in \mathscr{U}$ and $\alpha \neq 0$, there exists $V \in \mathscr{U}$ such that $\alpha V \subset U$.*

To prove this we observe by (2) and induction that, if $U \in \mathcal{U}$ and n is any positive integer, there exists $V \in \mathcal{U}$ such that $2^n V \subset U$. Now consider U and α in (5). Choose n so that $|\alpha| < 2^n$, and $V \in \mathcal{U}$ so that $2^n V \subset U$. Since U is balanced it is readily seen that $\alpha V \subset U$. Thus (5) is proved.

Now suppose that α_0 and x_0 are fixed. We shall show that, if $U \in \mathcal{U}$ there is a $V \in \mathcal{U}$ and an $\epsilon > 0$ such that $\alpha x \in \alpha_0 x_0 + U$ if $|\alpha - \alpha_0| < \epsilon$ and $x \in x_0 + V$. Choose W in $\mathcal{U}$ so that $W + W + W \subset U$ [by double use of (2) and observing that $0 \in W$]. If $\alpha_0 = 0$, let $V = W$; if $\alpha_0 \neq 0$, choose $W_1 \in \mathcal{U}$ so that $\alpha_0 W_1 \subset W$ [by (5)], and then choose $V \in \mathcal{U}$ so that $V \subset W \cap W_1$. By (1) there is some $\epsilon > 0$ such that $\beta x_0 \in V$ if $|\beta| < \epsilon$. We can assume that $\epsilon \leqslant 1$. Now suppose that $|\alpha - \alpha_0| < \epsilon$ and $x \in x_0 + V$, or $x - x_0 \in V$. We know that $(\alpha - \alpha_0)x_0 \in V$, and $(\alpha - \alpha_0)(x - x_0) \in V$ as a result of the fact that V is balanced. Finally, $\alpha_0(x - x_0) \in \alpha_0 V \subset W$. Consequently, from

$$\alpha x - \alpha_0 x_0 = (\alpha - \alpha_0)x_0 + \alpha_0(x - x_0) + (\alpha - \alpha_0)(x - x_0)$$

we see that $\alpha x - \alpha_0 x_0 \in W + W + W \subset U$. This completes the proof that multiplication is continuous.

The uniqueness assertion in Theorem 3.3–F follows from Theorem 3.3–A and the fact that a base determines its topology uniquely.

The definition of a topological linear space makes no provision for separation axioms. Thus, the space need not be a T_1-space. Consider, for instance, a linear space X having more than one point, with $\emptyset$ and X as the only open sets. The next theorem shows, however, that a T_1-topological linear space must necessarily be a T_3-space, and so, in particular, a Hausdorff space.

Theorem 3.3–G. *A topological linear space is regular. Hence, if it is a T_1-space, it is also a Hausdorff space.*

PROOF. Regularity is defined in § 2.3. We choose a base $\mathcal{U}$ at 0 having the property 3 in Theorem 3.3–E. Suppose U, $V \in \mathcal{U}$ and $V + V \subset U$, and suppose $x \in \bar{V}$. Now $x + (-1)\cdot V$ is a neighborhood of x, and therefore $\{x + (-1)\cdot V\} \cap V \neq \emptyset$. Suppose $y \in \{x + (-1)\cdot V\} \cap V$, and write $y = x - z$, $z \in V$. Then $x = y + z \in V + V \subset U$. This proves $\bar{V} \subset U$. The complete proof of the theorem now follows from Theorem 3.3–A and the fact that the closure of a set $x + V$ is $x + \bar{V}$.

It is also true that a topological linear T_1-space is completely regular (definition in § 2.3). For a proof of a somewhat more general result which is easily rephrased for the present situation, see Weil, **1**, page 13.

Definition. Two topological linear spaces X and Y are called *topologically isomorphic*, or *linearly homeomorphic*, if there exists a linear operator T which establishes a homeomorphic mapping of X onto all of Y.

Are all n-dimensional topological linear spaces with the same scalar field linearly homeomorphic to one another (and so in particular to $l^2(n)$)? Not necessarily so, for, if X is n-dimensional, we may give it the topology in which the only open sets are $\emptyset$ and X, and then it is certainly not homeomorphic to $l^2(n)$. However, this anomalous situation is avoided if we deal with T_1-spaces.

Theorem 3.3–H. *All n-dimensional topological linear T_1-spaces with the same scalar field are linearly homeomorphic.*

This is a generalization of Theorem 3.12–A. The proof is left for the problems at the end of the section.

Since the concept of distance is not available to us in discussing the general theory of topological linear spaces, the concept of a bounded set must be defined in a manner different from that for metric spaces.

Definition. In a topological linear space a set S is said to be *bounded* if for each neighborhood U of 0 there is some scalar α such that $S \subset \alpha U$.

In view of the fact that there exists a base at 0 composed of balanced neighborhoods, we may say "some positive scalar α" instead of "some scalar α," and the modified definition is equivalent to the original.

It is clear that a subset of a bounded set is bounded.

In the particular case in which the space is a normed linear space, the foregoing definition is easily seen to be equivalent to the definition of boundedness in terms of the norm; that is, S is bounded if and only if there is some positive β such that $\|x\| \leq \beta$ for each x in S.

The concept of boundedness is important in connection with the question of whether a topological linear space is normable (see Theorem 3.41–D).

In this book no extensive use is made of the general theory of topological linear spaces. Apart from normed linear spaces, the topological linear spaces studied in this book are mainly those which arise from the introduction of "weak topologies" into normed linear spaces. However, the general theory is a subject of growing interest, and much research is being done on it.

PROBLEMS

1. To prove Theorem 3.3–H it suffices to show that an n-dimensional topological linear T_1-space X is linearly homeomorphic to $l^2(n)$. If we write $x = \xi_1 x_1 + \cdots + \xi_n x_n$, where $x_1, \cdots, x_n$ is a basis for X, the main difficulty lies in proving that the point $(\xi_1, \cdots, \xi_n)$ of $l^2(n)$ depends continuously on x. The following suggestions should enable the reader to make the proof. If

$\epsilon > 0$ is given, let K be the set of points in $l^2(n)$ for which $|\xi_1|^2 + \cdots + |\xi_n|^2 = \epsilon^2$, and let S be the corresponding set in X. The fact that X is a T_1-space is used in proving that S is closed. It is possible to choose a balanced neighborhood of 0 in X, say U, such that $U \cap S = \emptyset$. One can then show that $x \in U$ implies $|\xi_1|^2 + \cdots + |\xi_n|^2 < \epsilon^2$.

2. If X and Y are topological linear spaces and X is a finite-dimensional T_1-space, every linear operator on X into Y is continuous.

3. A set T is said to absorb a set S if there is some $r > 0$ such that $S \subset \alpha T$ when $|\alpha| \geqslant r$. Show that a set S is bounded if and only if each balanced neighborhood of 0 absorbs S.

4. A set S is bounded if and only if it has the following property: If $\{\alpha_n\}$ is a sequence of scalars such that $\alpha_n \to 0$ and if $\{x_n\}$ is a sequence of elements of S, the sequence $\{\alpha_n x_n\}$ is convergent to 0.

5. Let X be the linear space of real-valued functions x which are defined and continuous on the real axis. Corresponding to each $\epsilon > 0$ and each compact set S of real numbers, let

$$U_{\epsilon, S} = \{x : |x(s)| < \epsilon \text{ if } s \in S\}.$$

As ϵ and S vary, the family $\mathscr{U}$ of the sets $U_{\epsilon, S}$ satisfies the conditions 1–4 of Theorem 3.3–F, and hence $\mathscr{U}$ is a base at 0 for a topology on X, making X into a topological linear space. Is it a T_1-topology?

6. On occasion it is useful to have an alternative to Theorem 3.3–F for the construction of a topological linear space. Here is such an alternative. Let X be a linear space, and let $\mathscr{W}$ be a nonempty family of nonempty sets with the following properties:

a. Each W in $\mathscr{W}$ is balanced and absorbing.
b. If W_1 and W_2 are in $\mathscr{W}$, there exists $W_3 \in \mathscr{W}$ such that $W_3 \subset W_1 \cap W_2$.
c. If $W_1 \in \mathscr{W}$, there exists $W_2 \in \mathscr{W}$ such that $W_2 + W_2 \subset W_1$.

Let a set S in X be defined as an open set if to each $x \in S$ there corresponds some $W \in \mathscr{W}$ such that $x + W \subset S$. With this family of open sets, X is a topological linear space. Verification of this assertion is left to the reader.

If we compare this situation with that in Theorem 3.3–F, we observe that property 4 of Theorem 3.3–F is not assumed as a property of the family $\mathscr{W}$. Otherwise the situations are the same. As a consequence, we are not able to assert that the members of $\mathscr{W}$ are open sets. Suppose, for instance, that X is the Euclidean plane and that $\mathscr{W}$ is the family of sets W_ϵ of the form

$$W_\epsilon = \{(\xi_1, \xi_2) : \xi_1{}^2 + \xi_2{}^2 \leqslant \epsilon^2\},$$

where ϵ can vary over all positive numbers. The resulting topology for X is the usual topology of the plane.

7. Consider $L^p(a, b)$, where $0 < p < 1$. We write

$$\|x\| = \left(\int_a^b |x(s)|^p \, ds\right)^{1/p},$$

just as when $p \geqslant 1$. With $0 < p < 1$ $\|x\|$ is not a norm, for the triangular inequality fails to hold. Instead, we have the inequality

$$\|x + y\| \leqslant \max\{2\|x\|, 2\|y\|\}.$$

However, let $W_\epsilon = \{x : \|x\| \leqslant \epsilon\}$. The family $\mathscr{W}$ of all such sets W_ϵ, for all $\epsilon > 0$, satisfies conditions *a–c* of 6. Hence L^p becomes a topological linear space if open sets are defined as follows: S is open if to each $x \in S$ there corresponds some $\epsilon > 0$ such that $y \in S$ whenever $\|y - x\| \leqslant \epsilon$.

8. If X and Y are topological linear spaces, a linear operator T on X into Y is continuous at all points of X if it is continuous at some point of X.

9. There is a generalization of the concept of completeness for a topological linear space. Since the topology may not be that of a metric space, it turns out to be desirable to use Moore-Smith convergence instead of sequential convergence in defining completeness. We rely on the terminology concerning *directed sets* and *nets* used by Kelley, **1**, Chapter 2, in connection with Moore-Smith convergence. Suppose X is a topological linear space. A net f in X (i.e., a function with range in X and domain some directed set D) is called a *Cauchy net* if for each neighborhood U of 0 in X there is some $d_0 \in D$ such that $f(d_1) - f(d_2) \in U$ if $d_0 < d_1$ and $d_0 < d_2$. The net is said to converge to x if for each neighborhood U of 0 there is some $d_0 \in D$ such that $f(d) - x \in U$ when $d_0 < d$. If a net converges to a point, it is a Cauchy net. If every Cauchy net in X converges to a point in X, we say that X is complete. An alternative approach to completeness is through the concept of *Cauchy filters* (see Bourbaki, **2**, Chapter II).

a. If X and Y are topological linear spaces which are topologically isomorphic, they are both complete if one is.

b. For a normed linear space the present definition of completeness is equivalent to the notion of completeness for metric spaces, as defined in § 2.41.

c. Let M be a closed subspace of the topological linear space X. If X is complete, so is M. If X is a T_1-space and if M is complete, it is closed in X, regardless of whether X is complete. For the proof it is necessary to know that, if $S \subset X$, a point x is in $\bar{S}$ if and only if there is a net in S converging to x. Also, when X is T_1, it is also T_2, and then a net in X cannot converge to two different points.

d. If X is a topological linear space which is locally compact and T_1, it is finite dimensional. We sketch the argument. Using the local compactness, we obtain a balanced neighborhood of 0, say V, such that $\bar{V}$ is balanced and compact. Then, for each neighborhood W of 0 we obtain as follows an $\alpha \neq 0$ such that $\alpha\bar{V} \subset W$. Choose a balanced neighborhood W_1 of 0 such that $W_1 + W_1 \subset W$. Choose points $x_1, \cdots, x_k \in \bar{V}$ so that $x_1 + W_1, \cdots, x_k + W_1$ cover $\bar{V}$. Choose $\alpha_i \neq 0$ so that $\alpha_i x_i \in W_1$. Then take α so that $|\alpha| \leqslant 1$ and $0 < |\alpha| \leqslant \min|\alpha_i|$. One can show $\alpha\bar{V} \subset W$. Since $\bar{V}$ is compact we can choose $y_1, \cdots, y_n \in \bar{V}$ so that $y_1 + {}^1/_2\bar{V}, \cdots, y_n + {}^1/_2\bar{V}$ cover $\bar{V}$. Let M be the linear manifold generated by $y_1, \cdots, y_n$. Evidently $\bar{V} \subset M + {}^1/_2\bar{V}$. Since X is T_1, M is topologically isomorphic to $l^2(m)$ for some $m \leqslant n$. Hence M is complete and therefore closed in X. Finally, $M = X$. For, assume $x_0 \in X - M$. Let A be the set of all nonzero scalars α such that $(x_0 + \alpha\bar{V}) \cap M \neq \emptyset$. Then A is not empty, as a result of the fact that $\bar{V}$ is absorbing. Let $\delta = \inf|\alpha|$ for $\alpha \in A$. The following argument shows that $\delta > 0$. Since $X - M$ is open there exists a neighborhood W of 0 such that $x_0 + W \subset X - M$. There exists a nonzero scalar α_1 such that $\alpha_1\bar{V} \subset W$. Then, using the fact that $\bar{V}$ is balanced, we show that α is not in A if $0 < |\alpha| \leqslant |\alpha_1|$, and so $\delta \geqslant |\alpha_1|$. Now choose $\alpha \in A$ so that $|\alpha| < 2\delta$ and choose $x \in (x_0 + \alpha\bar{V}) \cap M$. Then $(x - x_0)/\alpha \in \bar{V} \subset M + {}^1/_2\bar{V}$, and so we can write $x - \alpha y_0 = x_0 + (\alpha/2)v$ with $y_0 \in M$ and $v \in \bar{V}$. We conclude that $(\alpha/2) \in A$, which is a contradiction. Hence $M = X$ and X is finite dimensional.

e. Every finite-dimensional subspace of X is closed if X is a T_1-space.

10. Let M be a closed subspace of the topological linear space X. With the quotient space X/M and the canonical mapping ϕ of X onto X/M as defined in § 3.14, define a set S in X/M to be open if $\phi^{-1}(S)$ is open in X. This makes X/M a topological linear T_1-space. If the topology of X arises from a norm, this topology for X/M is the same as that generated by the norm defined in Theorem 3.14–A.

3.4 Convex Sets

Let X be a real linear space. If x_1 and x_2 are distinct points of X, the set of all points $(1 - \alpha)x_1 + \alpha x_2$ for which $0 \leqslant \alpha \leqslant 1$ is called the *line segment* joining x_1 and x_2, or simply the line segment x_1x_2.

Definition. A set S in X is called *convex* if S contains the line segment x_1x_2 whenever x_1 and x_2 are distinct points of S. This is equivalent to the requirement that, if $x_1, x_2 \in S$, then also $\alpha_1 x_1 + \alpha_2 x_2 \in S$ whenever α_1 and α_2 are positive numbers such that $\alpha_1 + \alpha_2 = 1$.

The empty set and a set consisting of one point are convex, since the conditions of the definition are vacuously satisfied by such sets. Notice that a linear manifold is convex and that the intersection of a family of convex sets is a convex set.

In a normed linear space an open sphere is a convex set. Likewise the closed sphere $\{x : \|x - x_0\| \leqslant r\}$ is convex.

The notion of a convex set is independent of topology, but convex sets play an important part in the theory of topological linear spaces. The first five theorems of this section make no use of topology. The remaining theorems involve topological ideas. The student who is interested in normed linear spaces, to the exclusion of more general types of topological linear spaces, may ignore Theorem 3.4–H.

If X is a complex linear space, it can also be regarded as a real linear space, and it is this point of view which is taken when we define the concept of a convex set in a complex linear space.

Theorem 3.4–A. *Let X be a linear space (real or complex). Let S and T be convex sets in X, and let α and β be any scalars. Then the set $\alpha S + \beta T$ is convex.*

We omit the proof, which is immediate.

Theorem 3.4–B. *Let S be a convex set in X, and let α, β be positive. Then $\alpha S + \beta S = (\alpha + \beta)S$.*

PROOF. It is evident that $(\alpha + \beta)S \subset \alpha S + \beta S$, regardless of convexity. If S is convex, and $y = \alpha x_1 + \beta x_2$, where $x_1, x_2 \in S$, then

$$\frac{1}{\alpha + \beta} y = \frac{\alpha}{\alpha + \beta} x_1 + \frac{\beta}{\alpha + \beta} x_2 \in S,$$

since

$$\frac{\alpha}{\alpha + \beta} + \frac{\beta}{\alpha + \beta} = 1.$$

Thus $y \in (\alpha + \beta)S$, and $\alpha S + \beta S \subset (\alpha + \beta)S$. This completes the proof.

Definition. If S is any set in X, the *convex hull* of S is defined to be the intersection of all convex sets which contain S. We sometimes denote it by S_c.

Theorem 3.4–C. *The convex hull of S consists of all points which are expressible in the form $\alpha_1 x_1 + \cdots + \alpha_n x_n$, where $x_1, \cdots, x_n$ are any points of S, $\alpha_k > 0$ for each k and $\sum_k \alpha_k = 1$. The index n is not fixed.*

PROOF. Let T be the set of points expressible in the manner described in the theorem. It is readily seen by the definition of convexity that T is convex. Since T clearly contains S, it follows that $S_c \subset T$. We shall show that T is contained in every convex set which contains S. We do this by induction on the number of points of S occurring in the representation of a point of T. Let W be a convex set containing S. If $x = \alpha_1 x_1$ is a point of T for which $n = 1$, it is clear that $\alpha_1 = 1$ and $x \in S$. Let us assume that a point of T is in W if it is represented in terms of $n - 1$ points of S. Then, let $x = \alpha_1 x_1 + \cdots + \alpha_n x_n$ be in T. Let $\beta = \alpha_1 + \cdots + \alpha_{n-1}$, $\beta_k = \alpha_k/\beta$ $(k = 1, \cdots, n - 1)$, $y = \beta_1 x_1 + \cdots + \beta_{n-1} x_{n-1}$. Then $y \in W$, by the induction hypothesis. But $x_n \in W$, $\alpha_n = 1 - \beta$, and $x = \beta y + (1 - \beta) x_n \in W$, since W is convex. This completes the induction. Thus $T \subset W$, and it follows that $T \subset S_c$. The proof is now complete.

We shall later on have occasion to deal with sets which are both balanced and convex. The concept of a balanced set involves considering all scalars α such that $|\alpha| \leq 1$. Accordingly, if X is a complex linear space, a set in X which is balanced when complex scalars are considered is also balanced when only real scalars are considered and X is viewed as a real linear space. But a set may be balanced from this latter point of view and *not* balanced when complex scalars are considered. To avoid confusion, we shall always understand that, when balanced sets are mentioned without explicit mention of the field of scalars under consideration, the sets are balanced relative to the field of scalars associated with the space.

Theorem 3.4–D. *A convex set in a real linear space is balanced if and only if it is symmetric.*

PROOF. Since every balanced set is symmetric, it suffices to show that a convex and symmetric set S is balanced. Suppose $x \in S$. Then $-x \in S$. Hence, since S is convex, $(1-\alpha)(-x) + \alpha x = (2\alpha - 1)x$ is in S if $0 \leqslant \alpha \leqslant 1$. Thus $\beta x \in S$ if $-1 \leqslant \beta \leqslant 1$, and S is balanced, by definition.

Let X be a linear space, either real or complex. If $S \subset X$, let S_b denote the balanced hull of S, and let S_c denote the convex hull of S.

Definition. The set $(S_b)_c$ is called the *balanced and convex hull* of S. We denote it simply by S_{bc}.

Theorem 3.4–E. *If X is a real or complex linear space, the set S_{bc} consists of all finite sums $\sum \alpha_k x_k$, where the x_k's are elements of S and $\sum |\alpha_k| \leqslant 1$. S_{bc} is the intersection of all sets which are both balanced and convex and contain S.*

PROOF. Let T be the set described in the theorem. We recall that S_b is the set of all elements αx, where $|\alpha| \leqslant 1$ and $x \in S$. If $x \in S_{bc}$, we know by Theorem 3.4–C that x is a finite sum $\sum \beta_k \alpha_k x_k$, where $x_k \in S$, $|\alpha_k| \leqslant 1$, $\beta_k > 0$, and $\sum \beta_k = 1$. It is thus clear that $S_{bc} \subset T$. On the other hand, suppose $x \in T$. Then $x = \sum \alpha_k x_k$, where $x_k \in S$, and $\sum |\alpha_k| \leqslant 1$. If $\sum |\alpha_k| = 0$ it is clear that $x = 0 \in S_b$ and so $x \in S_{bc}$. If $\sum |\alpha_k| = \alpha \neq 0$, let $\beta_k = |\alpha_k|/\alpha$. Then $\beta_k \geqslant 0$, $\sum \beta_k = 1$. We can write $\alpha_k = |\alpha_k| \gamma_k$, where γ_k is a scalar (real or complex) such that $|\gamma_k| = 1$. Let $y_k = \alpha \gamma_k x_k$. Then $\beta_k y_k = \alpha_k x_k$, and $x = \sum \beta_k y_k$. But $|\alpha \gamma_k| = \alpha \leqslant 1$, so $y_k \in S_b$. Thus $x \in S_{bc}$, by Theorem 3.4–C. We now know that $T \subset S_{bc}$ and therefore that $T = S_{bc}$. It is clear from the form of the elements of T that S_{bc} is balanced. It is of course convex. Verification of the last assertion in the theorem is left to the reader.

It may be remarked that $(S_c)_b$ is not always the same as $(S_b)_c$. Indeed, $(S_c)_b$ can fail to be convex.

Next we consider convex sets in a topological linear space.

Theorem 3.4–F. *Let S be a convex set in a topological linear space. Then the closure $\bar{S}$ is also convex.*

PROOF. Suppose $x_0, y_0 \in \bar{S}$ and $0 < \alpha < 1$. Consider the function $f(x, y) = (1 - \alpha)x + \alpha y$. It is continuous. Let U be any neighborhood of $(1 - \alpha)x_0 + \alpha y_0$. If we show that U contains points of S, it will follow that $(1 - \alpha)x_0 + \alpha y_0 \in \bar{S}$ and hence that $\bar{S}$ is convex. Now, since f is continuous, there exist neighborhoods V and W of x_0 and y_0 respectively such that $f(x, y) \in U$ if $x \in V$ and $y \in W$. Since $x_0 \in \bar{S}$, V contains some point $x_1 \in S$; likewise W contains some point $y_1 \in S$. Thus, since S is convex, $f(x_1, y_1) \in S$. But $f(x_1, y_1) \in U$ also. Thus the proof is complete.

If S is any set, the closure of the convex hull of S is called the *closed convex hull* of S. It is the intersection of all closed and convex sets which contain S. The sets S and $\bar{S}$ have the same closed convex hull.

Theorem 3.4–G. *If S is an open set, its convex hull S_c is also open.*

PROOF. Suppose $x \in S_c$, then x is a finite sum $\sum \alpha_k x_k$, where $x_k \in S$, $\alpha_k > 0$, and $\sum \alpha_k = 1$. Since S is open, there exist neighborhoods V_k of x_k such that $V_k \subset S$. Let T be the set $\sum \alpha_k V_k$, i.e., the set of all elements $\sum \alpha_k v_k$, where $v_k \in V_k$. Clearly $x \in T \subset S_c$, so all that is needed is to show that T is open. Now each $\alpha_k V_k$ is open, since $\alpha_k \neq 0$. Hence T will be open, by induction, once we know that $U + V$ is open if U and V are open. Now $x + V$ is open for each fixed x, and $U + V$ is the union of all $x + V$ as x varies over U. Thus $U + V$ is open.

The next theorem is used in the proof of Theorem 3.41–D. The result would be obvious and a formal statement unnecessary if we confined ourselves to normed linear spaces.

Theorem 3.4–H. *If X is a topological linear space and U is a convex neighborhood of* 0, *then U contains a balanced and convex neighborhood of* 0.

PROOF. We know from the proof of Theorem 3.3–D that U contains a balanced neighborhood V of 0. Let W be the convex hull of V. Then W is a neighborhood of 0, by Theorem 3.4–G. Since V is its own balanced hull, W is the balanced and convex hull of V and is therefore balanced (Theorem 3.4–E). Since U is convex, it follows that $W \subset U$. Thus W meets the requirements of the theorem. This argument works for both real and complex spaces. For real spaces we could obtain a satisfactory W by taking $W = U \cap (-U)$ (see Theorem 3.4–D).

PROBLEMS

1. Let S be a convex set in the topological linear space X. Suppose $x_0 \in \text{int}(S)$ and $y_0 \in \bar{S}$. Then, every point y expressible in the form $y = \alpha x_0 + (1 - \alpha)y_0$, $0 < \alpha < 1$, is an interior point of S.

2. Under the conditions on S and X in problem 1, the set int (S) is convex. If int $(S) \neq \emptyset$, we have $\overline{\text{int}(S)} = \bar{S}$ and int $(S) = $ int $(\bar{S})$.

3. Let $A_1, \cdots, A_n$ be compact convex sets in the topological linear space X. Let $A = A_1 \cup \cdots \cup A_n$. Then A_c is compact. Begin by showing that A_c is composed of all the points x which are expressible in the form $x = \alpha_1 x_1 + \cdots + \alpha_n x_n$, with $x_k \in A_k$, $0 \leqslant \alpha_k$, and $\alpha_1 + \cdots + \alpha_n = 1$. Then, let P be the compact set in $l^2(n)$ composed of all points $(\alpha_1, \cdots, \alpha_n)$ with $\alpha_k \geqslant 0$ and $\alpha_1 + \cdots + \alpha_n = 1$. Observe that A_c is the image in X of the product set $P \times A_1 \times \cdots \times A_n$ by the mapping $(\alpha_1, \cdots, \alpha_n, x_1, \cdots, x_n) \rightarrow \alpha_1 x_1 + \cdots + \alpha_n x_n$.

4. In a normed linear space X the convex hull of a precompact set S is precompact. The argument depends on Theorem 2.41–G and problem 3. If $\epsilon > 0$, choose $x_1, \cdots, x_n$ in X so that S is contained in the union B of the open spheres of radii $\epsilon/2$ with centers $x_1, \cdots, x_n$. Let A be the finite set $x_1, \cdots, x_n$. Show that B_c is contained in the union of the open spheres of radii $\epsilon/2$ with centers in A_c. Then observe that A_c is compact and so deduce that S_c is precompact.

5. In a Banach space X the closed convex hull of a compact set is compact. Use problem 4.

3.41 Minkowski Functionals

In this section we assume that X is a real linear space and that K is a convex set in X. We assume further that K is an absorbing set and that $0 \in K$.

Definition. For each $x \in X$ let A_x be the set of those real α such that $\alpha > 0$ and $x \in \alpha K$. Since K is absorbing, A_x is not empty. We then define

$$p(x) = \inf A_x. \tag{3.41–1}$$

The functional p is called the Minkowski functional of the set K. Evidently $0 \leqslant p(x) < \infty$. This functional is useful in dealing analytically with K. We shall see in § 3.7 that there is a close relation between Minkowski functionals and seminorms, which are used in connection with the topology of certain topological linear spaces.

If X is a normed linear space and $K = \{x : \|x\| < r\}$, where r is fixed and positive, it is easy to see that $p(x) = \|x\|/r$. If K is a closed and bounded convex set in Euclidean space of n dimensions, if 0 is an interior point of K and $x \neq 0$, we can think of $p(x)$ as follows: Let y be that unique positive multiple of x in which the ray from 0 through x intersects the boundary of K. Then $x = p(x)y$, and $p(x)$ is the ratio of the distance $0x$ to the distance $0y$.

Theorem 3.41–A. *Let K be a convex, absorbing set which contains 0. The Minkowski functional of K has the properties*

$$p(0) = 0, \qquad p(\lambda x) = \lambda p(x) \qquad \textit{if } \lambda > 0 \tag{3.41–2}$$

$$p(x+y) \leqslant p(x) + p(y). \tag{3.41–3}$$

If K is balanced, then, for any scalar λ,

$$p(\lambda x) = |\lambda| p(x). \tag{3.41–4}$$

This last relation holds even for complex λ if X happens to be a complex space and K is balanced in the complex sense.

PROOF. If x and y are given and α, β are arbitrary elements of A_x, A_y respectively, we have $x + y \in \alpha K + \beta K$. This last set is $(\alpha + \beta)K$, by Theorem 3.4–B, and so $\alpha + \beta \in A_{x+y}$. It follows that $p(x + y) \leqslant \alpha + \beta$. Owing to the arbitrariness of α and β, (3.41–3) now follows. The relation $p(0) = 0$ is evident, since $0 \in K$. To prove the other relation in (3.41–2) suppose $\lambda > 0$ and $x \in X$. Take any $\alpha \in A_x$. Then $x \in \alpha K$, $\lambda x \in \lambda\alpha K$, and so $\lambda\alpha \in A_{\lambda x}$, whence $p(\lambda x) \leqslant \lambda\alpha$. By the arbitrariness of α we conclude that $p(\lambda x) \leqslant \lambda p(x)$. Now replace x by λx and λ by λ^{-1}, thus obtaining $p(x) \leqslant \lambda^{-1} p(\lambda x)$, or $\lambda p(x) \leqslant p(\lambda x)$. Consequently $p(\lambda x) = \lambda p(x)$. If K is balanced, if $\lambda \neq 0$ and $x \in X$, suppose $\alpha \in A_x$. Then $x \in \alpha K$, $(\lambda/|\lambda|\alpha)x \in K$, $\lambda x \in |\lambda|\alpha K$, $p(\lambda x) \leqslant |\lambda|\alpha$, and so $p(\lambda x) \leqslant |\lambda| p(x)$. The reverse inequality is obtained by the same device as before, and so (3.41–4) is proved.

As a result of (3.41–3) we note the following inequalities, which will be used later:

$$-p(y - x) \leqslant p(x) - p(y) \leqslant p(x - y). \tag{3.41–5}$$

Theorem 3.41–B. *Let K be a convex, absorbing set which contains 0. If p is the Minkowski functional of K, then $p(x) \leqslant 1$ whenever $x \in K$, and $p(x) < 1$ implies that $x \in K$.*

PROOF. If $x \in K$, then $1 \in A_x$ and so $p(x) \leqslant 1$. If $p(x) < 1$, there exists $\alpha \in A_x$ such that $0 < \alpha < 1$. Now, since 0 and $\alpha^{-1}x$ are in K and K is convex, it follows that $\alpha \cdot \alpha^{-1}x + (1 - \alpha)\cdot 0 = x \in K$.

Next we add the assumption that X is a topological linear space.

Theorem 3.41–C. *Let X be a topological linear space, and let p be the Minkowski functional of a set K in X, where K is convex, absorbing, and contains 0. Let $K_1 = \{x : p(x) < 1\}$, $K_2 = \{x : p(x) \leqslant 1\}$. Then*

a. $\operatorname{int}(K) \subset K_1 \subset K \subset K_2 \subset \bar{K}$.
b. $K = K_1$ *if K is open.*
c. $K = K_2$ *if K is closed.*
d. *If p is continuous,* $K_1 = \operatorname{int}(K)$ *and* $K_2 = \bar{K}$.
e. *p is continuous if and only if* $0 \in \operatorname{int}(K)$.
f. *If K is bounded and X is a T_1-space,* $p(x) = 0$ *implies* $x = 0$.

PROOF. Suppose $x \in \operatorname{int}(K)$. Then $\alpha x \in K$ when α is sufficiently near 1, and hence we can find some $\alpha > 1$ such that $\alpha x \in K$. Then $\alpha^{-1} \in A_x$ and $p(x) \leqslant \alpha^{-1} < 1$, so $x \in K_1$. We know $K_1 \subset K \subset K_2$ from Theorem 3.41–B. To prove $K_2 \subset \bar{K}$ it suffices to consider an x for which $p(x) = 1$. Then $\alpha x \in K_1 \subset K$ if $0 < \alpha < 1$. Letting $\alpha \to 1$ we see that $\alpha x \to x$, whence $x \in \bar{K}$. We have now finished with (*a*); (*b*) and (*c*) are immediate consequences. If p is continuous, K_1 is open by Theorem 2.12–A; hence, in view of (*a*), $K_1 = \operatorname{int}(K)$; it also follows that K_2 is closed and so $K_2 = \bar{K}$. We observe that $0 \in K_1$. Then $0 \in \operatorname{int}(K)$ if p is continuous, by (*d*). On the other hand, if $0 \in \operatorname{int}(K)$, let U be a neighborhood of 0 such that $U \subset \operatorname{int}(K)$. Then $p(x) < 1$ if $x \in U$. Now, if $\epsilon > 0$, then ϵU is a neighborhood of 0, and we have $p(x) < \epsilon$ if $x \in \epsilon U$, because $\epsilon^{-1}x \in U$ and $p(\epsilon^{-1}x) = \epsilon^{-1}p(x) < 1$. This shows that p is continuous at 0. The inequalities (3.41–5) then show that p is continuous at all other points (recall that every neighborhood of 0 contains a symmetric neighborhood of 0). Finally, suppose that X is a T_1-space and that K is bounded. If $x \neq 0$, there exists a balanced neighborhood of 0 which does not contain x. Let U be this neighborhood. Since K is bounded, $K \subset \beta U$ for some $\beta > 0$. Now suppose $\alpha \in A_x$. Then $x \in \alpha K$ and so $x/\alpha \in \beta U$, or $(1/\alpha\beta)x \in U$. Since U is balanced and x is not in U, it follows that $\alpha\beta > 1$, for $\alpha\beta \leqslant 1$ would imply $x = \alpha\beta(1/\alpha\beta)x \in U$. But $\alpha > 1/\beta$ implies $p(x) \geqslant 1/\beta$, and so $p(x) > 0$. This finishes the proof of the theorem.

Definition. A topological linear space X is said to be *normable* if there exists (i.e., if it is possible to define) a norm on X such that the resulting topology for X as a normed linear space is the same as the given topology for X.

Theorem 3.41–D. *A topological linear space X is normable if and only if* (1) *X is a T_1-space and* (2) *there exists in X a convex and bounded neighborhood of* 0.

PROOF. The conditions are evidently necessary, for a metric space is a T_1-space, and the set $\{x : \|x\| < 1\}$ in a normed linear space is both convex and bounded. Suppose, on the other hand, that X satisfies conditions 1 and 2 of the theorem. Let U be a bounded and convex neighborhood of 0. There exists a neighborhood V of 0 such that $V \subset U$ and V is balanced as well as convex and bounded (Theorem 3.4–H). Let p be the Minkowski functional of V. By Theorem 3.41–A we see that p has all the properties of a norm except possibly for the property that $p(x) \neq 0$ if $x \neq 0$, and this property follows by Theorem 3.41–C, part *f*. Thus p is a norm on X

We now complete the proof by showing that the topology of X is the same as the topology defined on X by the norm p. We know by Theorem 3.41–C, part b, that $V = \{x:p(x) < 1\}$. If $\alpha > 0$, $\alpha V = \{x:p(x) < \alpha\}$. It suffices to prove that the family of sets $\{\alpha V\}$ (α the parameter) is a base at 0 in the given topology of X. Suppose U is any neighborhood of 0. There exists a balanced neighborhood of 0, say W, such that $W \subset U$. Since V is bounded, there exists some $\epsilon > 0$ such that $V \subset \epsilon W$. Thus $\epsilon^{-1}V \subset U$. This shows that $\{\alpha V\}$ is a base at 0, and the proof is complete.

3.5 Linear Varieties

Throughout this section X denotes a linear space; except where specific mention is made of the nature of the scalar field, it may be either the real or the complex field.

The considerations of this section are geometric in character. If we suppose that X is Euclidean space of three dimensions, a linear variety is either a point, a line, a plane, or the whole space. The general notion of linear variety requires merely a linear space for its setting, with no need of considerations of topology or finite dimensionality. The most interesting linear varieties are those known as *hyperplanes*. A hyperplane M through 0 in X can be characterized in two ways: (1) It is the set of elements x such that $x'(x) = 0$, where x' is some nonzero linear functional on X; (2) it has the property that X can be expressed as the direct sum of M and a one-dimensional subspace.

Definition. A set $M \subset X$ is called a *linear variety* if $M = x_0 + M_0$, where x_0 is a fixed vector and M_0 is a subspace of X.

We call M a translation of M_0. Note that $x_0 \in M$, because $0 \in M_0$. It is easy to see that, if $M = x_0 + M_0$ and $x_1 \in M$, then also $M = x_1 + M_0$. This follows at once from the fact that, if x_1 and x_2 are in M, then $x_1 - x_2 \in M_0$.

A linear variety is a convex set. A set consisting of a single point is a linear variety, for (0) is a subspace.

A subspace X_0 of X will be called *maximal* if it is not all of X and if there exists no subspace X_1 of X such that $X_0 \neq X_1$, $X_1 \neq X$ and $X_0 \subset X_1$. A linear variety which results from the translation of a maximal subspace is called a *hyperplane*. In particular, a hyperplane containing 0 is the same thing as a maximal subspace. All of the hyperplanes obtained by translating a particular one are said to be parallel to this one.

At this point we recall the notation X^f for the linear space of all linear functionals on X (see § 1.6).

Theorem 3.5–A. *If $x' \in X^f$, $x' \neq 0$, and α is any fixed scalar, the set $M = \{x : x'(x) = \alpha\}$ is a hyperplane. It contains the origin 0 if and only if $\alpha = 0$. To each hyperplane M containing 0 corresponds an $x' \in X^f$ such that $x' \neq 0$ and $M = \{x : x'(x) = 0\}$.*

PROOF. Consider the first assertion. Since $x' \neq 0$, there exists x_1 such that $x'(x_1) = \beta \neq 0$. Let $x_0 = (\alpha/\beta)x_1$. Then $x'(x_0) = \alpha$, so $x_0 \in M$. Let $M_0 = -x_0 + M$. It is quickly verifiable that $M_0 = \{x : x'(x) = 0\}$. Clearly M_0 is a subspace, and so M is a linear variety. Now $M_0 \neq X$. If $y \in X - M_0$, every element of X can be written as the sum of an element of M_0 and a multiple of y. In fact, if $x \in X$ and

$$z = x - \frac{x'(x)}{x'(y)}\,y,$$

we see that $x'(z) = 0$, so that x has the required form. If we now suppose that M_1 is a subspace of X for which $M_0 \subset M_1$ and $M_1 \neq M_0$, we can choose $y \in M_1 - M_0$. The foregoing argument then shows that $X \subset M_1$, so that $M_1 = X$. This shows that M_0 is maximal and that M is a hyperplane. Evidently M contains 0 if and only if $\alpha = 0$.

Now suppose that M is any hyperplane containing 0. Theorem 1.71–B shows that there exists an $x' \neq 0$ such that $M \subset \{x : x'(x) = 0\}$. But then we must have $M = \{x : x'(x) = 0\}$; otherwise M would not be maximal. This completes the proof.

In the course of the proof of Theorem 3.5–A we obtained the following result:

Theorem 3.5–B. *Suppose $x' \in X^f$, $x' \neq 0$, and $M = \{x : x'(x) = 0\}$. Then, if $x_0 \in X - M$, every $x \in X$ can be expressed in the form*

$$x = \frac{x'(x)}{x'(x_0)}\,x_0 + m, \qquad m \in M. \tag{3.5–1}$$

This is a convenient place at which to insert the following useful theorem:

Theorem 3.5–C. *Suppose that $x_1', \cdots, x_n'$ are linearly independent elements of X^f, and let $y' \in X^f$ be such that $y'(x) = 0$ whenever $x_1'(x) = \cdots = x_n'(x) = 0$. Then y' is a linear combination of $x_1', \cdots, x_n'$.*

PROOF. For the case $n = 1$ we get the proof directly from (3.5–1). If $x_1' \neq 0$, and if $y'(x) = 0$ whenever $x_1'(x) = 0$, we have

$$y' = \frac{y'(x_0)}{x_1'(x_0)}\,x_1',$$

where x_0 is any element for which $x_1'(x_0) \neq 0$. We now proceed inductively, assuming the theorem true for a set of $n - 1$ linearly independent elements of X^f. It then follows, since $x_1', \cdots, x_n'$ are linearly independent, that for each $k (k = 1, \cdots, n)$ there exists an x_k such that $x_k'(x_k) \neq 0$ and $x_j'(x_k) = 0$ if $j \neq k$. Because of homogeneity we may arrange to have $x_k'(x_k) = 1$. Now let

$$y = x - \sum_{k=1}^{n} x_k'(x)x_k.$$

We then see that $x_j'(y) = 0$ for each j; hence also $y'(y) = 0$, by our assumption. But then

$$y'(x) = \sum_{k=1}^{n} x_k'(x)y'(x_k);$$

this is equivalent to

$$y' = \sum_{k=1}^{n} y'(x_k)x_k',$$

and the theorem is proved.

Theorem 3.5–D. *Let X be a topological linear space. The closure of a linear variety in X is a linear variety.*

PROOF. Suppose first that M is a linear subspace. Now, the following general proposition is true: If f is a continuous function which maps one topological space into another and S is any set in the first space, $f(\bar{S}) \subset \overline{f(S)}$. This is easily proved, directly from the definition of continuity. Now consider f to be the function which maps $X \times X$ into X, the value of f at (x, y) being $x + y$. We know that f is continuous. Hence $f(\bar{M} \times \bar{M}) \subset \overline{f(M \times M)}$. But, since M is a subspace, $f(M \times M) \subset M$. Thus $f(\bar{M} \times \bar{M}) \subset \bar{M}$. This shows that $x + y \in \bar{M}$ if $x \in \bar{M}$ and $y \in \bar{M}$. A similar argument, using the continuity of αx, shows that $\alpha x \in \bar{M}$ if $x \in \bar{M}$. Thus $\bar{M}$ is a subspace. A translation is a homeomorphism; therefore the closure of any linear variety is a linear variety.

Theorem 3.5–E. *Let X be a topological linear space and M a hyperplane containing 0. Then, either M is a closed set, or it is everywhere dense in X. Suppose $M = \{x : x'(x) = 0\}$, where x' is a fixed element of X^f. Then M is closed if and only if x' is continuous.*

PROOF. We have $M \subset \bar{M} \subset X$, and $\bar{M}$ is a subspace. Since M is maximal, we have either $M = \bar{M}$ or $\bar{M} = X$. This proves the first assertion. If x' is continuous, M is closed, for it is the inverse image (by x') of a single point in the scalar field. Suppose, on the other hand, that M is closed but that x' is not continuous. Then it is not continuous at 0, and there is some $\epsilon > 0$ such that every neighborhood U of 0 contains some point for which $|x'(x)| \geqslant \epsilon$. Now let $L = \{x : x'(x) = \epsilon/2\}$. The set L is obtained from M by translation. In fact, if $x_0 \in X - M$, L is the set of all elements of the form $[\epsilon/2x'(x_0)]x_0 + m$, $m \in M$, as we see by (3.5–1). Therefore L is closed. Hence there exists a neighborhood of 0 which does not meet L. We can choose a balanced neighborhood U of this kind, by Theorem 3.3–E. Now U must contain a point x_1 such that $|x'(x_1)| \geqslant \epsilon$. Let $\alpha = [\epsilon/2x'(x_1)]$. Then $\alpha x_1 \in U$, for U is balanced. But $x'(\alpha x_1) = \epsilon/2$, so that $\alpha x_1 \in L$, and we have a contradiction. Therefore x' is continuous.

We remark, for later use, that, if a hyperplane is closed, all the translations of it are likewise closed and that, if it is not closed, each of these translations is everywhere dense in X.

3.6 Convex Sets and Hyperplanes

In this section we assume throughout that X is a real linear space; all scalars which occur are real.

Our object in this section is to discuss certain geometrical relationships between convex sets and hyperplanes. Some of the theorems have important uses in connection with applications.

If $M = \{x : x'(x) = \alpha\}$ is a hyperplane ($x' \neq 0$, α fixed), each of the sets

$$\{x : x'(x) < \alpha\}, \qquad \{x : x'(x) > \alpha\}, \qquad \{x : x'(x) \leqslant \alpha\}, \qquad \{x : x'(x) \geqslant \alpha\}$$

is convex. Each of these sets is called a *half space* determined by M. If X is a topological linear space and x' is continuous, the first two of these half spaces are open and the last two are closed.

We say that a set S *lies on one side* of M if S lies entirely in one of the four half spaces determined by M. If S lies on one side of M and does not intersect M, we say that S is *strictly* on one side of M.

Theorem 3.6–A. *In order that a convex set S lie strictly on one side of a hyperplane M, it is necessary and sufficient that the intersection $S \cap M$ be empty.*

PROOF. The condition is necessary, by definition. Let $M = \{x : x'(x) = \alpha\}$, suppose that x_1 and x_2 are in S and that $x'(x_1) < \alpha < x'(x_2)$. Now $\lambda x_1 + (1 - \lambda)x_2$ lies in S if $0 \leqslant \lambda \leqslant 1$, and $x'(\lambda x_1 + (1 - \lambda)x_2) = \lambda x'(x_1) + (1 - \lambda)x'(x_2)$ is a continuous function of λ which has the value $x'(x_2)$ at $\lambda = 0$ and the value $x'(x_1)$ at $\lambda = 1$. It must then have the value α at some λ for which $0 < \lambda < 1$. But then $\lambda x_1 + (1 - \lambda)x_2 \in S \cap M$. Hence $S \cap M = \emptyset$ implies that S lies strictly on one side of M.

Theorem 3.6–B. *Let K be a convex, absorbing set which contains* 0, *and let p be its Minkowski functional* (see § 3.41). *Let M be a hyperplane which does not intersect K. There exists $x' \in X^f$ such that $M = \{x : x'(x) = 1\}$, and for this x' we have*

$$-p(-x) \leqslant x'(x) \leqslant p(x) \tag{3.6–1}$$

if $x \in X$.

PROOF. Any M which does not contain 0 can be represented in the form $M = \{x : x'(x) = \alpha\}$ where $\alpha \neq 0$. This follows readily from Theorem 3.5–A. We can then, on multiplying x' by a suitable constant if necessary, assume that $\alpha = 1$. Since $0 \in K$ and $x'(0) = 0$, we see (Theorem 3.6–A) that K lies strictly on one side of M and in the half space $\{x : x'(x) < 1\}$. Thus $x'(x) < 1$ if $x \in K$. Now consider any $x \in X$, and suppose $\alpha \in A_x$ (notation as in § 3.41). Then $0 < \alpha$, $\alpha^{-1}x \in K$, $x'(\alpha^{-1}x) < 1$, and $x'(x) < \alpha$. But then $x'(x) \leqslant p(x)$, by the definition of $p(x)$. We also have $x'(-x) \leqslant p(-x)$, whence $-p(-x) \leqslant -x'(-x) = x'(x)$. This completes the proof.

Theorem 3.6–C. *Let X be a topological linear space and M a hyperplane. Suppose that S is a set having at least one interior point and lying on one side of M. Then M is closed; the interior of S lies strictly on one side of M, and the closure $\bar{S}$ also lies on one side of M.*

PROOF. Let $M = \{x : x'(x) = \alpha\}$. We can suppose that S lies in the half space $\{x : x'(x) \leqslant \alpha\}$ (otherwise replace x' by $-x'$ and α by $-\alpha$). Now M must be closed, for otherwise the hyperplane $\{x : x'(x) = \alpha + 1\}$ would be everywhere dense and would, therefore, intersect S (since S has an interior point). Consequently x' is continuous (Theorem 3.5–E). But then the set $\{x : x'(x) < \alpha\}$ is open, and it is the interior of the closed set $\{x : x'(x) \leqslant \alpha\}$, for it is easy to see that x is *not* an interior point of the latter set if $x'(x) = \alpha$ (if $\alpha \neq 0$ merely consider λx for λ sufficiently near 1; otherwise use a preliminary translation). The assertions about the interior and closure of S now follow at once.

Theorem 3.6–D. *Let K be a convex absorbing set which contains* 0, *and let p be its Minkowski functional. Let L be a linear variety which does not intersect K. Then there exists a hyperplane M with equation $x'(x) = 1$ such that M contains L and $x'(x) \leqslant p(x)$ for every x. The set K lies in the half space $\{x : x'(x) \leqslant 1\}$. If X is a topological linear space and K has interior points, x' is continuous and M is closed. In this case the closure of K lies in the half space $\{x : x'(x) \leqslant 1\}$, and the interior of K lies in the open half space $\{x : x'(x) < 1\}$. In particular, if K is open, K does not intersect the closed hyperplane M.*

PROOF. Choose $x_0 \in L$ and let $L_0 = -x_0 + L$. Observe that x_0 is not in L_0, for, if it were, we should have $L = L_0$; then $0 \in L \cap K$, whereas $L \cap K$ is empty. Let X_0 be the set of all elements of the form $x = \alpha x_0 + y$, $y \in L_0$. X_0 is a subspace, and L_0 is a maximal subspace of X_0. For, suppose L_1 is a subspace such that $L_0 \subset L_1 \subset X_0$ and $L_0 \neq L_1$. Choose $x_1 \in L_1 - L_0$. Then $x_1 = \alpha_1 x_0 + y_1$, where $y_1 \in L_0$ and $\alpha_1 \neq 0$. Now, if $x = \alpha x_0 + y \in X_0$, we can write $x = (\alpha/\alpha_1)x_1 + z$, where $z = -(\alpha/\alpha_1)y_1 + y$. Since $z \in L_0$, we see that $x \in L_1$. Thus $L_1 = X_0$. This shows the maximality of L_0 in X_0. Now consider $K \cap X_0$. It is a convex and absorbing set in X_0; it contains 0, and its Minkowski functional is the restriction of p to X_0. Since $K \cap X_0$ does not intersect L, we know by Theorem 3.6–B that there exists $g \in (X_0)^f$ such that $L = \{x : x \in X_0,\ g(x) = 1\}$ and $g(x) \leqslant p(x)$ if $x \in X_0$. By Theorem 1.71–D there exists $x' \in X^f$ such that $x'(x) = g(x)$ if $x \in X_0$ and $x'(x) \leqslant p(x)$ if $x \in X$. The hyperplane $M = \{x : x'(x) = 1\}$ evidently contains L. Theorem 3.41–B shows that $x'(x) \leqslant 1$ if $x \in K$. If we assume that X is a topological linear space, the last assertions of the theorem follow from Theorem 3.6–C.

Theorem 3.6–E. *Let K be a nonempty open convex set in a topological linear space X, and let L be a linear variety which does not intersect K. Then, there exists a closed hyperplane M which contains L and is such that K lies strictly on one side of M.*

PROOF. The theorem is true if it is true for the case in which $0 \in K$, as we see by making suitable translations. But, if $0 \in K$, K is absorbing, and the theorem results at once from Theorem 3.6–D.

PROBLEMS

A general X referred to in these problems is a real topological linear space.

1. If K_1 and K_2 are nonempty, nonintersecting convex sets and if K_1 is open, there exists a closed hyperplane M such that K_1 is in one of the two closed half spaces determined by M and K_2 is in the other. If K_2 is also open, M can be chosen so that K_1 and K_2 are strictly on opposite sides of M. Argument: $K = K_1 + (-1)K_2$ is convex, nonempty, open, and does not contain 0. Hence there exists a closed hyperplane through 0 and not intersecting K. A suitable parallel hyperplane can be chosen for M.

2. If $S \subset X$, a *support* of S is a hyperplane M such that S lies on one side of M and $S \cap M \neq \emptyset$. If $x_0 \in S \cap M$, we say that S is supported by M at x_0. A closed convex set with a nonvacuous interior is called a *convex body*. Show that a support of a convex body is closed and that the body is supported at every boundary point. (See problem 2, § 3.4.)

3. Let K be a convex body in X and not all of X. Consider the closed half spaces containing K, and determined by the supports of K (see problem 2). The intersection of all these half spaces is K.

4. There exists a nonzero continuous linear functional x' defined on X if and only if in X there is at least one convex set which is open and contains 0 but is not all of X. For the "if" part, use Theorem 3.6–E, taking L to be a point not in the convex set. M. M. Day (see Day, **1**) has shown that there exists no nonzero continuous linear functional on $L^p(a, b)$ if $0 < p < 1$. For the topology of $L^p(a, b)$ see § 3.3, problem 7.

3.7 Seminorms

The concept of a seminorm plays an important role in the study of topological linear spaces. Seminorms are analytical devices for dealing with convex sets of a special sort. Seminorms are useful in describing the topology of a linear space if it is of the special type known as a locally convex topological linear space (see § 3.8).

Definition. Let X be a linear space (real or complex), and let p be a function with real values defined on X such that

(*a*) $$p(x_1 + x_2) \leqslant p(x_1) + p(x_2) \qquad \text{if } x_1, x_2 \in X,$$

(*b*) $$p(\alpha x) = |\alpha| p(x) \qquad \text{if } x \in X \text{ and } \alpha \text{ is any scalar.}$$

Then p is called a *seminorm* on X.

If p has the further property that $p(x) \neq 0$ if $x \neq 0$, p is a *norm* (see § 3.1). As with a norm, the properties of p imply the further properties

$$p(0) = 0, \tag{3.7–1}$$

$$p(x) \geqslant 0, \tag{3.7–2}$$

and

$$|p(x_1) - p(x_2)| \leqslant p(x_1 - x_2). \tag{3.7–3}$$

As we see by Theorem 3.41–A, if p is the Minkowski functional of a convex, balanced, and absorbing set K, then p is a seminorm. The following theorem shows that the converse is also true.

Theorem 3.7–A. *Let X be a linear space and p a seminorm on X. Let V be the set $\{x : p(x) < 1\}$. Then V is convex, balanced, and absorbing, and p is the Minkowski functional of V. If we suppose in addition that X is a topological linear space and that p is continuous, the set V is open.*

PROOF. Suppose x_1, $x_2 \in V$ and $0 \leqslant \alpha \leqslant 1$. Then $p(x_1) < 1$, $p(x_2) < 1$, and

$$p[(1 - \alpha)x_1 + \alpha x_2] \leqslant (1 - \alpha)p(x_1) + \alpha p(x_2) < 1,$$

by the properties a and b of p. Thus V is convex. The fact that V is balanced and absorbing follows from the homogeneity property b of p. Let q be the Minkowski functional of V. By definition, this means that $q(x)$ is the infimum of positive numbers α such that $x \in \alpha V$. But $\alpha > 0$ and $x \in \alpha V$ are together equivalent to $\alpha > 0$ and $p(x) < \alpha$. It is therefore clear that $p(x) = q(x)$. The final assertion of the theorem, about V being open, is true as a consequence of Theorem 2.12–A.

Before considering the next theorem we shall make some general observations about linear functionals on complex linear spaces. Suppose that X is a complex linear space. Then, we can also regard X as a real linear space simply by restricting our attention to real scalars. Let us denote by X_r the real linear space obtained from X by adopting this point of view. Now consider a linear functional $x' \in X^f$. If $x \in X$ let us denote the real and imaginary parts of $x'(x)$ by $x_1'(x)$ and $x_2'(x)$, respectively, so that $x'(x) = x_1'(x) + ix_2'(x)$. Since $x'(ix) = ix'(x)$, we have

$$x_1'(ix) + ix_2'(ix) = ix_1'(x) - x_2'(x),$$

whence $x_2'(x) = - x_1'(ix)$, and so

$$x'(x) = x_1'(x) - ix_1'(ix). \tag{3.7–4}$$

It is readily verified that $x_1' \in (X_r)^f$, i.e., that x_1' is a *real*-valued linear functional on the *real* linear space X_r. Conversely, if we start with *any* element $x_1' \in (X_r)^f$ and define x' by (3.7–4), it is easy to verify that $x'(\alpha x) = \alpha x'(x)$ for all *complex* scalars and that $x' \in X^f$.

Theorem 3.7–B. *Let X be a linear space, and let p be a seminorm defined on X. Let M be a linear manifold in X and f a linear functional defined on M such that $|f(x)| \leqslant p(x)$ if $x \in M$. Then, there exists a linear functional F defined on X such that $|F(x)| \leqslant p(x)$, if $x \in X$, and $F(x) = f(x)$ if $x \in M$.*

PROOF. Observe that the seminorm p is a sublinear functional (defined just preceding Theorem 1.71–D). Since $p(-x) = p(x)$, it may be noted that, for a linear F, $F(x) \leqslant p(x)$ for all x implies $|F(x)| \leqslant p(x)$. For $-F(x) = F(-x) \leqslant p(-x) = p(x)$, or $-p(x) \leqslant F(x)$. The two inequalities yield $|F(x)| \leqslant p(x)$.

For the case in which X is a real space, the present theorem is a direct application of Theorem 1.71–D. If X is a complex space, we proceed as follows: Denote X by X_r when we consider it as a linear space with real scalars; likewise with M and M_r. We write $f(x) = f_1(x) - if_1(ix)$, with $f_1(x)$ the real part of $f(x)$. Then f_1 is a real linear functional on M_r [by the discussion accompanying (3.7–4)]. Now, $|f_1(x)| \leqslant |f(x)| \leqslant p(x)$ if $x \in M$. Hence, by what we already know for the real case, there exists a real linear functional F_1 on X_r such that $F_1(x) = f_1(x)$ if $x \in M_r$ and $|F_1(x)| \leqslant p(x)$ if $x \in X$. We define $F(x) = F_1(x) - iF_1(ix)$. Then F is a linear functional on the complex space X, and $F(x) = f(x)$ if $x \in M$. For any given x let $F(x) = re^{i\theta}$ $(r \geqslant 0)$. Then $F(e^{-i\theta}x) = e^{-i\theta}F(x) = r$. Since this is a real value, $F(e^{-i\theta}x) = F_1(e^{-i\theta}x)$. Then $|F(x)| = F_1(e^{-i\theta}x) \leqslant p(e^{-i\theta}x) = p(x)$. This completes the proof.

Theorem 3.7–C. *Let X be a topological linear space, x_0 a point of X, and p a seminorm on X which is continuous. Then, there exists a continuous linear functional F defined on X, such that $F(x_0) = p(x_0)$ and $|F(x)| \leqslant p(x)$ if $x \in X$.*

PROOF. Let M be the set of all elements αx_0, and define f on M by $f(\alpha x_0) = \alpha p(x_0)$. Then f is linear on M and $|f(\alpha x_0)| = p(\alpha x_0)$. Hence, by Theorem 3.7–B, there exists a linear functional F defined on X such that F is an extension of f and $|F(x)| \leqslant p(x)$. This inequality shows that F is continuous at 0 (and hence continuous at all points) because p is continuous. Clearly $F(x_0) = f(x_0) = p(x_0)$.

Theorem 3.7–D. *Let X be a topological linear space, and let V be a subset of X which is closed, balanced, convex, and of which 0 is an interior point. Let x_0 be a point on the boundary of V. Then, there exists a continuous linear functional F defined on X such that $F(x_0) = 1$ and $|F(x)| \leqslant 1$ if $x \in V$.*

PROOF. Let p be the Minkowski functional of V (see § 3.41). The properties of V assure us that p is a seminorm (by Theorem 3.41–A), and by Theorem 3.41–C we see that $V = \{x : p(x) \leqslant 1\}$, that p is continuous, and that the interior of V is the set $\{x : p(x) < 1\}$. Consequently $p(x_0) = 1$, since x_0 is on the boundary of V. The conclusions now follow from Theorem 3.7–C.

One of the important uses of this theorem is in the proof of Theorem 4.7–C.

3.8 Locally Convex Spaces

Among topological linear spaces which are not normed linear spaces, there is a class of spaces for which theoretical developments along the lines drawn for normed linear spaces seem to be more satisfactory than is possible without restriction of attention to this class. This class of spaces is characterized by the existence of a base at 0 composed of convex sets. The spaces are said to be locally convex. The main reason for singling out locally convex spaces for special attention lies in the fact that for such spaces we can prove certain important theorems about the existence of continuous linear functionals. In this section we discuss briefly a few of the main features of locally convex spaces.

Definition. A topological linear space X is called *locally convex* if every neighborhood of 0 contains a convex neighborhood of 0.

If X is locally convex and $\mathscr{U}$ is the family of all balanced and convex neighborhoods of 0, this family is a base at 0 (see Theorem 3.4–H). It is easily verified that, if $U \in \mathscr{U}$, then $\alpha U \in \mathscr{U}$ when $\alpha \neq 0$; also, there exists $V \in \mathscr{U}$ such that $V + V \subset U$ (use Theorem 3.4–B).

One way of constructing a locally convex topology is described in the following theorem.

Theorem 3.8–A. *Suppose that X is a linear space and that $\mathscr{U}$ is a nonempty family of nonempty subsets of X with the properties:*

1. *Each member of $\mathscr{U}$ is balanced, convex, and absorbing.*
2. *If $U \in \mathscr{U}$, there exists some α such that $0 < \alpha \leqslant {}^1/_2$ and $\alpha U \in \mathscr{U}$.*
3. *If U_1 and U_2 are in $\mathscr{U}$, there exists $U_3 \in \mathscr{U}$ such that $U_3 \subset U_1 \cap U_2$.*
4. *If $U \in \mathscr{U}$ and $x \in U$, there exists $V \in \mathscr{U}$ such that $x + V \subset U$.*

Then there is a unique topology for X such that X is a locally convex topological linear space with $\mathscr{U}$ as a base at 0.

PROOF. If we compare this theorem with Theorem 3.3–F, we see that it suffices to prove that, if $U \in \mathscr{U}$, there exists $V \in \mathscr{U}$ such that $V + V \subset U$. With U given we select α as in (2), and let $V = \alpha U$. Then $V + V = \alpha U + \alpha U = 2\alpha U \subset U$ as a result of the fact that $2\alpha \leqslant 1$ and that U is convex and balanced.

Next we show how seminorms may be used to define a locally convex topology.

Theorem 3.8–B. *Let P be a nonempty family of seminorms defined on the linear space X. For each $p \in P$ let $V(p)$ be the set $\{x : p(x) < 1\}$. Let $\mathscr{U}$ be the family of all finite intersections*

$$r_1 V(p_1) \cap r_2 V(p_2) \cap \cdots \cap r_n V(p_n), \qquad r_k > 0, \qquad p_k \in P.$$

Then $\mathscr{U}$ satisfies the conditions of Theorem 3.8–A.

PROOF. The sets $V(p)$ are balanced, convex, and absorbing (Theorem 3.7–A). The members of $\mathscr{U}$ also have these properties. Verification of conditions 2 and 3 in Theorem 3.8–A is left to the reader. As for (4), if $U = r_1 V(p_1) \cap \cdots \cap r_n V(p_n)$ and $x \in U$, it suffices to take $V = \alpha_1 V(p_1) \cap \cdots \cap \alpha_n V(p_n)$, where $\alpha_k = r_k - p_k(x)$.

Observe that $rV(p) = \{x : p(x) < r\}$.

Definition. The topology for X determined by taking the family $\mathscr{U}$ of Theorem 3.8–B as a base at 0 is called the topology generated by the family P of seminorms.

Observe that, if P contains just one element p and if p is a norm, $\mathscr{U}$ is the family of all spheres $\{x:p(x) < r\}$, and the topology for X generated by P is the topology of X as a normed linear space with norm p.

If X is a locally convex space and P is the family of all continuous seminorms on X, the topology of X is exactly the topology generated by P. The fact that P is not empty follows from the results of § 3.41 (see especially Theorem 3.41–C, part *e*).

If $p_1, \cdots, p_n$ are seminorms on X and if $\alpha_1, \cdots, \alpha_n$ are positive, let us define

$$p(x) = \max \{\alpha_1 p_1(x), \cdots, \alpha_n p_n(x)\}.$$

Then p also is a seminorm. If $p_1, \cdots, p_n$ are continuous, so is p. These observations are useful in connection with the formulation of the continuity of a linear operator by means of conditions involving seminorms.

Theorem 3.8–C. *Let X and Y be locally convex topological linear spaces, and let Q be a family of seminorms which generates the topology of Y. Let T be a linear operator on X into Y. Then T is continuous on X if and only if to each $q \in Q$ there corresponds some continuous seminorm p on X such that $q(Tx) \leqslant p(x)$ for each $x \in X$.*

PROOF. We prove the "only if" assertion and leave the rest of the proof to the reader. Let T be continuous, and suppose $q \in Q$. Then $\{y:q(y) < 1\}$ is a neighborhood of 0 in Y. Hence there exists a set $V = r_1 V(p_1) \cap \cdots \cap r_n V(p_n)$ (in the notation of Theorem 3.8–B), determined by continuous seminorms $p_1, \cdots, p_n$ on X and positive numbers $r_1, \cdots, r_n$, such that $q(Tx) < 1$ if $x \in V$. Then

$$p(x) = \max_k \frac{1}{r_k} p_k(x)$$

defines a continuous seminorm p. If x is arbitrary in X and α is chosen so that $p(x) < \alpha$, it follows that $x/\alpha \in V$ and hence $q(Tx) < \alpha$. Since α can be as close to $p(x)$ as we please, we infer that $q(Tx) \leqslant p(x)$.

The reader will note that the condition for continuity in the foregoing theorem specializes to the condition $\|Tx\| \leqslant M\|x\|$ of Theorem 3.1–A in case the space is a normed linear space.

We come now to the theorems about the existence of continuous linear functionals in certain situations.

Theorem 3.8–D. *Let X be a locally convex topological linear space. Let M be a linear manifold in X, and let f be a continuous linear functional on M. Then, there exists a continuous linear functional F on X which is an extension of f.*

PROOF. Since f is continuous and X is locally convex, there exists a balanced and convex neighborhood of 0, say V, such that $|f(x)| < 1$ if $x \in M \cap V$. Let p be the Minkowski functional of V. Then p is a seminorm, it is continuous, and $V = \{x : p(x) < 1\}$ (Theorems 3.41–A and 3.41–C). For any $x \in M$ choose α so that $\alpha > p(x)$. Then $p(x/\alpha) < 1$ and so $|f(x/\alpha)| < 1$, or $|f(x)| < \alpha$. By letting $\alpha \to p(x)$ we see that $|f(x)| \leqslant p(x)$. Applying Theorem 3.7–B, we obtain a linear functional F on X which is an extension of f and such that $|F(x)| \leqslant p(x)$ for every x. This inequality and Theorem 3.8–C show that F is continuous, so the proof is complete.

Theorem 3.8–E. *Let X be a locally convex topological linear space. Let X_0 be a closed subspace of X, and suppose $x_1 \in X - X_0$. Then, there exists a continuous linear functional x' on X such that $x'(x_1) = 1$ and $x'(x) = 0$ if $x \in X_0$.*

PROOF. Consider the subspace M generated by x_1 and X_0, and represent elements of M uniquely in the form $m = \alpha x_1 + x$, $x \in X_0$. The set $x_1 + X_0$ (i.e., the set of elements of M for which $\alpha = 1$) is closed and does not contain 0. Hence there exists a balanced convex neighborhood of 0, which we denote by V, such that V does not intersect $x_1 + X_0$. Let p be the Minkowski functional of V; p is a continuous seminorm and $V = \{x : p(x) < 1\}$. We then have $p(x_1 + x) \geqslant 1$ if $x \in X_0$. Let us define a functional m' on M by setting $m'(m) = \alpha$. We assert that $|m'(m)| \leqslant p(m)$. If $\alpha = 0$, this is clear, and if $\alpha \neq 0$, we have $p(m) = p(\alpha x_1 + x) = |\alpha| p(x_1 + \alpha^{-1}x) \geqslant |\alpha| = |m'(m)|$, because $\alpha^{-1}x \in X_0$. Since p is continuous it now follows that m' is continuous. Observe that $m'(m) = 0$ if $m \in X_0$ and that $m'(x_1) = 1$. Applying Theorem 3.8–D we finish the proof.

Theorem 3.8–F. *Let X be a locally convex topological linear space which is a T_1-space. Then, if $x_0 \neq 0$, there exists a continuous linear functional x' on X such that $x'(x_0) = 1$.*

PROOF. We can take $X_0 = (0)$ and apply Theorem 3.8–E; X_0 will be closed, since X is a T_1-space.

The next theorem gives a simple criterion for a space to be a T_1-space.

Theorem 3.8–G. *If X is a locally convex topological linear space with* *topology generated by a family P of seminorms, then X is a T_1-space if and only if to each $x \neq 0$ corresponds $p \in P$ such that $p(x) \neq 0$.*

The proof is left to the reader.

The main reference to locally convex spaces in this book comes in connection with the study of "weak" topologies; one aspect of this subject is introduced in the following example, and other aspects are considered in § 3.81.

Example 1. Let T be any nonempty set. Let X be a linear space composed of elements which are real-valued (or complex-valued) functions x defined on T. X need not comprise *all* functions defined on T. For example, X might consist of all bounded functions or of all functions meeting some prescribed condition, provided that $x + y$ and αx meet the condition whenever x and y do. As a family of seminorms take

$$p_t(x) = |x(t)|$$

for each $t \in T$. This family of seminorms generates a locally convex topology for X. With this topology X is a T_1-space, by Theorem 3.8–G. We note that $x_n \to x$ (convergence in the sense of this topology) means that $x_n(t) \to x(t)$ (convergence in the usual sense for scalars) for each $t \in T$.

There is an aspect of this topology for X which is interesting and significant, namely, the following: For a given $t \in T$ consider $x(t)$ as a function of x, with x varying over X. This function is continuous on X in the topology just defined. For, if $x_0 \in X$ and $\epsilon > 0$ and U is the neighborhood of 0 defined by $U = \{x : p_t(x) < \epsilon\}$, we see that $|x(t) - x_0(t)| < \epsilon$ means the same thing as $x \in x_0 + U$. Suppose for a moment that we consider any topology for X which is such that $x(t)$ is continuous as a function of x, no matter how t is chosen in T. Then, if $\epsilon > 0$, the set $\{x : |x(t)| < \epsilon\}$ is open, for it is the inverse image of an open set of scalars. But

$$\{x : |x(t)| < \epsilon\} = \{x : p_t(x) < \epsilon\}.$$

Hence the given topology for X contains among its open sets the base at 0 in the topology for X generated by the family of seminorms. This shows that the latter topology is the weakest topology for X with respect to which $x(t)$ is continuous as a function of x for each t. It is called the *weak topology of X as a space of functions.* One specific example of T and X is that in which T is the set of positive integers and X is the space of *all* sequences $x = \{\xi_1, \xi_2, \cdots\}$. The weak topology of X arises from the demand that each ξ_n be a continuous function of x.

We conclude this section with an example of a different sort.

Example 2. Suppose c is either a positive real number or $+\infty$. Let X be the linear space of all functions x of the complex variable t representable as a power series

$$x(t) = \sum_{0}^{\infty} \xi_n t^n$$

which is convergent if $|t| < c$. The coefficients ξ_n are to be complex. For each R such that $0 < R < c$ let

$$p_R(x) = \sum_{0}^{\infty} |\xi_n| R^n.$$

It is readily evident that p_R is a norm on X. The family of these norms, obtained by varying R, generates a locally convex topology for X and makes X a T_1-space. Convergence in this topology is equivalent to uniform convergence on each compact subset of the set $\{t: |t| < c\}$. This is easily verified if we recall that when $0 < b < c$ and $M_b = \max |x(t)|$ for $|t| = b$, then

$$|\xi_n| \leqslant \frac{M_b}{b^n}.$$

It may also be verified that a set S in this space X is bounded, in the sense defined in § 3.3, if and only if the family of functions forming the set S is uniformly bounded on each compact subset of the set $\{t: |t| < c\}$. It is a well-known theorem (due to Montel) of classical function theory that from every infinite sequence in such a family of functions there can be selected a subsequence which converges uniformly on each compact subset of the domain on which the family is defined. This leads (by means of the fact that X is metrizable and that when metrized it is complete) to the conclusion that the space under consideration has the property that all of its closed and bounded subsets are compact. Since the space is not finite dimensional, the situation is in strong contrast with the situation for normed linear spaces (see Theorem 3.12–F). For information about metrizability, see § 3.9, especially Theorem 3.9–B.

PROBLEMS

1. Let P be the family of all seminorms on the linear space X, where $X \neq (0)$. This family is nonempty, and the topology it generates is the strongest locally convex topology on X. Moreover, it is a T_1-topology. Suggestion: Given $x_0 \neq 0$, choose a Hamel basis for X of which x_0 is a member. For any $x \in X$ let $p(x)$ be the absolute value of the coefficient of x_0 in the representation of x in terms of the Hamel basis.

2. If S is a set in the locally convex topological linear space X, S is bounded if and only if each continuous seminorm is bounded on S. Instead of *all* continuous seminorms it is sufficient to consider those in a family which generates the topology.

3. For the next problem we require the following result: Let A and B be disjoint sets in the topological linear space X (not necessarily locally convex). Assume A compact and B closed. Then, there exists a neighborhood of 0, say V, such that $A + V$ and $B + V$ are disjoint. Outline of argument: Assume $(A + V) \cap (B + V) \neq \emptyset$, no matter how V is chosen (as a neighborhood of 0). If V is symmetric, this implies that $A \cap (B + V + V) \neq \emptyset$, for we can choose $a \in A, b \in B, v_1 \in V, v_2 \in V$ so that $a + v_1 = b + v_2$, whence $a \in A \cap (B + V + V)$. Show that, if $V_1, \cdots, V_n$ are symmetric neighborhoods of 0, $\bigcap_{i=1}^{n} [A \cap (B + V_i + V_i)] \neq \emptyset$, and deduce the existence of $x_0 \in A \cap \bigcap_{V} \overline{[A \cap (B + V + V)]}$. Hence, for any symmetric V, $x_0 + V$ must intersect $A \cap (B + V + V)$. From this deduce that $x_0 \in B$, a contradiction.

4. Let X be a real locally convex topological linear space. Let K_1, K_2 be nonempty, nonintersecting convex sets in X, with K_1 closed and K_2 compact. Then, there exists a closed hyperplane M such that K_1 and K_2 lie strictly on opposite sides of M. Argument: Let V be a convex neighborhood of 0. Then $K_1 + V$ and $K_2 + V$ are open and convex. By problem 3 we may choose V so that $(K_1 + V) \cap (K_2 + V) = \emptyset$. Now apply the result of problem 1, § 3.6.

5. If K is a closed convex set in the real locally convex topological linear space X, then K is the intersection of all the closed half spaces which contain it. Can we also assert this for open half spaces?

6. Let K be a nonempty compact convex set in the real locally convex topological linear T_1-space. Consider the closed half spaces which contain K and which are determined by the supports of K (see problem 2, § 3.6). *The intersection of these half spaces is K.* It suffices to show that, if x_0 is a point not in K, there exists a support M of K such that x_0 and K are not in the same one of the two closed half spaces determined by M. Choose (x_0) and K as the K_1 and K_2 of problem 4, and obtain a separating closed hyperplane, say with equation $x'(x) = \alpha$. By considering the values of x' on K, a suitable support of K may be found as a hyperplane $x'(x) = \beta$.

3.81 Weak Topologies for Linear Spaces. Duality

In this section we consider briefly the basic notions about a class of topologies, called *weak topologies*, for a linear space. These notions are developed further in connection with normed linear spaces, in Chapter 4.

Suppose that X is any linear space (real or complex), and let F be a nonempty set in X^f. In what follows immediately hereafter we use A, A_1, A_2, etc. to denote nonempty finite subsets of F. For any such A and any $\epsilon > 0$ let

$$U(A; \epsilon) = \{x : |x'(x)| < \epsilon \quad \text{if } x' \in A\}.$$

It is no trouble to verify that $0 \in U(A; \epsilon)$, that $U(A; \epsilon)$ is balanced, convex, and absorbing, that $U(A_1 \cup A_2; \min [\epsilon_1, \epsilon_2]) \subset U(A_1; \epsilon_1) \cap U(A_2; \epsilon_2)$, and that $\alpha U(A; \epsilon) = U(A; |\alpha| \epsilon)$ if $\alpha \neq 0$. If $x \in U(A; \epsilon)$, let $\delta = \sup_{x' \in A} |x'(x)|$, so that $0 \leqslant \delta < \epsilon$. Then $x + U(A; \epsilon - \delta) \subset U(A; \epsilon)$. It follows by Theorem 3.8–A that the family of all sets $U(A; \epsilon)$ is a base at 0 for a uniquely defined locally convex topology on X. We denote this topology by $\mathscr{T}(X, F)$. If p is the Minkowski functional of $U(A; \epsilon)$, it is a seminorm which is continuous on X in the topology just defined. It is easy to demonstrate that

$$p(x) = \frac{1}{\epsilon} \sup_{x' \in A} |x'(x)|.$$

The following theorem is of basic importance to an understanding of the nature of the topology $\mathscr{T}(X, F)$.

Theorem 3.81–A. *An element x' of X^f is continuous on X with the topology $\mathscr{T}(X, F)$ if and only if x' is in the linear manifold generated by F.*

PROOF. The proof that x' is continuous if $x' \in F$ is easy, and we leave it to the reader. The "if" assertion follows at once. Now assume that $y' \in X^f$ and that y' is continuous in the $\mathscr{T}(X, F)$ topology. By Theorem 3.8–C, with Y the space of scalars, we see that there exists some finite subset A of F and some $\epsilon > 0$ such that

$$|y'(x)| \leqslant \frac{1}{\epsilon} \sup_{x' \in A} |x'(x)|$$

for every x in X. It follows that $y'(x) = 0$ whenever $x'(x) = 0$ for every x' in A. If $x_1', \cdots, x_n'$ is a maximal linearly independent subset of A, it follows by Theorem 3.5–C that y' is a linear combination of $x_1', \cdots, x_n'$. This completes the proof.

If A is a set consisting of a single element x', then $U(A; \epsilon)$ is the inverse image, under x', of the set of scalars α for which $|\alpha| < \epsilon$. It follows easily from this remark and Theorem 3.81–A that $\mathscr{T}(X, F)$ is the weakest topology for X which makes it a topological linear space and which makes all the elements of F continuous on X. This is the reason why $\mathscr{T}(X, F)$ is called a weak topology for X.

It follows from the foregoing remarks that, if M is the linear manifold generated by F, the topologies $\mathscr{T}(X, F)$ and $\mathscr{T}(X, M)$ are the same. Also, if M_1 and M_2 are linear manifolds in X^f, with M_1 a proper subset of M_2, the topology $\mathscr{T}(X, M_2)$ is strictly stronger than $\mathscr{T}(X, M_1)$.

An explanation of the conditions under which $\mathscr{T}(X, F)$ makes X a T_1-space depends upon the concept of a *total* set of linear functionals.

Definition. A set $F \subset X^f$ is called total if to each $x \neq 0$ in X there corresponds some x' in F such that $x'(x) \neq 0$. Or, equivalently, F is total if $x'(x) = 0$ for each $x' \in F$ implies $x = 0$.

For example, if $X = C[a, b]$ and F is the set of functionals of the form $x'(x) = x(r)$, where $a \leqslant r \leqslant b$ and r is rational, then F is total.

Theorem 3.81–B. *The topology $\mathscr{T}(X, F)$ makes X a T_1-space if and only if F is total.*

This follows directly from Theorem 3.8–G.

When F is total, we may identify X with a certain space of functions defined on F. We identify x with the function whose value at x' in F is $x'(x)$. The topology $\mathscr{T}(X, F)$ is then the same as the weak topology described in Example 1, § 3.8, where T is taken to be F.

Suppose now that M is a subspace of X^f. If x is a fixed element of X and x' is a variable element of M, the expression $x'(x)$ defines an element of M^f, and the set G of such elements is a linear manifold and a total set in M^f. Hence the weak topology $\mathscr{T}(M, G)$ for M makes M a locally convex topological linear space and a T_1-space. If M itself is total as a subset of X^f, we can identify X and G (i.e., the correspondence between x and the element of G which it defines establishes an isomorphism of X and G as linear spaces). In this case we shall write $\mathscr{T}(M, X)$ in place of $\mathscr{T}(M, G)$. There is a kind of *duality* here: M is the class of all linear functionals on X which are continuous in the topology $\mathscr{T}(X, M)$, and X is identified with the class G of all the linear functionals on M which are continuous in the topology $\mathscr{T}(M, X)$.

It sometimes happens that we consider weak topologies for a linear space X which is already a locally convex topological linear space, say with a topology $\mathscr{T}_0$. Let M be the class of all those elements of X^f which are continuous on X with respect to the topology $\mathscr{T}_0$. As we have already noted, the topology $\mathscr{T}(X, M)$ is weaker than $\mathscr{T}_0$ (though not necessarily strictly weaker). We observe, as a consequence of Theorem 3.8–F, that, if X is a T_1-space, then M will be total.

Theorem 3.81–C. *Let X, $\mathscr{T}_0$, and M be as described in the foregoing paragraph (X need not be a T_1-space). Then, a linear manifold in X is closed relative to the topology $\mathscr{T}_0$ if and only if it is closed relative to the topology $\mathscr{T}(X, M)$.*

PROOF. Let L be the linear manifold in question. We prove that L is closed relative to $\mathscr{T}(X, M)$ if it is closed relative to $\mathscr{T}_0$. The other half of the proof is evident by taking the complement of L and using the fact that $\mathscr{T}(X, M)$ is weaker than $\mathscr{T}_0$. Suppose $x_0 \in X - L$. By

Theorem 3.8–E there exists $x' \in M$ such that $x'(x_0) = 1$ and $x'(x) = 0$ if $x \in L$. Suppose $0 < \epsilon < 1$. The set $\{x : |x'(x - x_0)| < \epsilon\}$ is a neighborhood of x_0 in the topology $\mathscr{T}(X, M)$. This neighborhood lies in $X - L$, and therefore $X - L$ is open relative to $\mathscr{T}(X, M)$. Then L is closed relative to $\mathscr{T}(X, M)$.

PROBLEMS

1. Let X be any linear space and M any subspace of X^f. A nonempty subset $S \subset X$ is bounded relative to the topology $\mathscr{T}(X, M)$ if and only if $\sup\limits_{x \in S} |x'(x)| < \infty$ for each $x' \in M$.

2. Let X be a locally convex topological linear T_1-space. Let M be the set of those $x' \in X^f$ which are continuous on X. Let V_0 be a balanced and convex neighborhood of 0 which is bounded relative to the topology $\mathscr{T}(X, M)$, and let p_0 be the Minkowski functional of V_0. Then p_0 is a norm. It can be shown that the topology generated by p_0 is the same as the given topology of X. See problem 8, § 4.4.

3. Let X be a complex linear space, and let X_r be the associated real linear space (as in § 3.7). If $x' \in X^f$ and $x'(x) = x_1'(x) - ix_1'(ix)$, where $x_1' \in (X_r)^f$ [see (3.7–4)], we may look upon x' as an element of $(X^f)_r$, and the correspondence $x' \leftrightarrow x_1'$ establishes an isomorphism of $(X^f)_r$ and $(X_r)^f$. Let M be a complex linear manifold in X^f, and let M_1 be the corresponding real linear manifold in $(X_r)^f$. Show that the topology $\mathscr{T}(X, M)$, as applied to X_r, coincides with $\mathscr{T}(X_r, M_1)$.

4. Let X be a locally convex topological T_1-space, and let S be a convex set in X. Let X' be the set of all the continuous linear functionals on X. Then, the closure of S with respect to the given topology of X is the same as the closure of S with respect to the weak topology $\mathscr{T}(X, X')$. Use Theorem 3.81–C and problem 5 in § 3.8.

3.9 Metric Linear Spaces

Suppose that X is a linear space and that there is a metric $d(x_1, x_2)$ defined on X. If the metric has the property

$$d(x_1 + y, x_2 + y) = d(x_1, x_2) \tag{3.9–1}$$

for every choice of x_1, x_2 and y, we call it an *invariant metric.* If the metric is invariant, the choice $y = -x_2$ shows that

$$d(x_1 - x_2, 0) = d(x_1, x_2). \tag{3.9–2}$$

Conversely, if (3.9–2) holds for all choices of x_1, x_2, the metric is invariant, as we see at once.

Definition. We shall say that X is a *metric linear space* if it is a topological linear space with topology derived from an invariant metric $d(x_1, x_2)$.

A normed linear space is a metric linear space, for the metric $d(x_1, x_2) = \|x_1 - x_2\|$ derived from a norm is certainly invariant. However, there are metric linear spaces in which the metric is not derived from a norm. In order to see the similarities and differences when comparing metric linear spaces with normed linear spaces, it is convenient to abbreviate $d(x, 0)$ by the notation $|x|$. We see from the properties of the invariant metric that

$$|x + y| \leqslant |x| + |y|, \tag{3.9–3}$$

$$|-x| = |x|, \tag{3.9–4}$$

$$|x| = 0 \quad \text{if and only if } x = 0. \tag{3.9–5}$$

There is nothing to guarantee that $|\alpha x| = |\alpha|\,|x|$, however, and so $|x|$ may not be a norm. An example will be given presently. Of course, $d(x, 0)$ is a continuous function of x and αx is a continuous function of α and x; therefore $|\alpha x|$ is a continuous function of α and x. It follows that

$$|\alpha_n x| \to 0 \quad \text{if } |\alpha_n| \to 0, \tag{3.9–6}$$

$$|\alpha x_n| \to 0 \quad \text{if } |x_n| \to 0, \tag{3.9–7}$$

$$|\alpha_n x_n| \to 0 \quad \text{if } |\alpha_n| \to 0 \text{ and } |x_n| \to 0. \tag{3.9–8}$$

The following theorem shows how a metric linear space may be characterized by specifying properties of $|x|$ rather than properties of a metric.

Theorem 3.9–A. *Let X be a linear space, and suppose $|x|$ is a real-valued function on X with the six properties* (3.9–3)–(3.9–8). *Make the definition $d(x_1, x_2) = |x_1 - x_2|$. Then d is an invariant metric on X, and, with this metric, X is a metric linear space.*

The proof is left to the reader. For continuity of scalar multiplication we observe that

$$|\alpha_n x_n - \alpha x| \leqslant |(\alpha_n - \alpha)x| + |\alpha(x_n - x)| + |(\alpha_n - \alpha)(x_n - x)|.$$

Some simple examples of nonnormed, but metric, linear spaces are based on the inequality

$$\frac{|A + B|}{1 + |A + B|} \leqslant \frac{|A|}{1 + |A|} + \frac{|B|}{1 + |B|}, \tag{3.9–9}$$

which is valid if A and B are arbitrary real or complex numbers. A proof of (3.9–9) follows readily from the observation that $x(1 + x)^{-1}$ increases as the real variable x increases, if $x > -1$ (check by computing the derivative).

Example 1. The space (s). The class of all sequences $x = \{\xi_n\}$ becomes a metric linear space if we define

$$\|x\| = \sum_{n=1}^{\infty} \mu_n \frac{|\xi_n|}{1 + |\xi_n|}, \tag{3.9–10}$$

where $\{u_n\}$ is any fixed sequence of positive numbers such that $\sum_{1}^{\infty} \mu_n$ is convergent. We denote this space by (s).

Convergence in (s) is merely componentwise convergence. That is, if $x^{(i)} = \{\xi_n^{(i)}\}$, then $\|x^{(i)} - x\| \to 0$ as $i \to \infty$ if and only if $\lim_{i\to\infty} \xi_n^{(i)} = \xi_n$ for each n. Proof is left to the reader. It follows that (s) is a complete metric space.

Example 2. *The space S.* Let $[a, b]$ be a finite closed interval of the real axis. Consider the family of all real-valued functions defined on $[a, b]$ which are measurable in the Lebesgue sense, but not necessarily bounded. By using almost-everywhere equality as an equivalence relation this family gives rise to a family of equivalence classes which we make into a linear space (see, e.g., the discussion of $\mathcal{L}^p$ and L^p in Example 7, § 1.2). If x is one of the equivalence classes, we define

$$\|x\| = \int_a^b \frac{|x(t)|}{1 + |x(t)|}\, dt, \tag{3.9–11}$$

where $x(t)$ denotes a representative function for x. We then obtain a metric linear space, which we denote by S. We can also consider the case in which the functions are complex-valued.

Convergence in S is interpretable in terms of convergence *in measure.* We recall the definition: $x_n(t)$ converges in measure to $x(t)$ if for each $\epsilon > 0$ the measure of the set $\{t: |x_n(t) - x(t)| \geqslant \epsilon\}$ converges to zero as $n \to \infty$. Let $E_n(\epsilon)$ be the set in question. By decomposing $[a, b]$ into the union of $E_n(\epsilon)$ and its complement we see from (3.9–11) that

$$\|x_n - x\| \leqslant m(E_n(\epsilon)) + \epsilon(b - a),$$

where m refers to Lebesgue measure. Also, since $\alpha(1 + \alpha)^{-1}$ is an increasing function of α when $\alpha > -1$, we see that

$$\|x_n - x\| \geqslant \int_{E_n(\epsilon)} \frac{|x_n(t) - x(t)|}{1 + |x_n(t) - x(t)|}\, dt \geqslant \frac{\epsilon}{1 + \epsilon} m(E_n(\epsilon)).$$

From the foregoing considerations it is clear that $|x_n - x| \to 0$ as $n \to \infty$ is equivalent to convergence in measure of $x_n(t)$ to $x(t)$. From a well-known fact about convergence in measure it then follows that S is a complete metric space (see, e.g., Halmos, **2**, Theorem E, page 93).

A complete metric linear space has been called a space of type (F) by Banach (Banach, **1**, Chapter III). The F is for Fréchet, who made important contributions in the early studies of linear spaces with a metric. Currently, however, the designation "a space of type (F)" is sometimes used with the additional assumption that the space is locally convex (see, e.g., Dieudonné, **2**, page 499).

We say that a topological linear space X is *metrizable* if its topology is metrizable, i.e., if there is a metric in X such that the metric topology is the same as the given topology of X. Now, in a metric space there is a countable base of neighborhoods at each point (e.g., the interiors of spheres of radii $1/n$ with centers at the point, $n = 1, 2, \cdots$). Moreover, a metric space is of course a T_1-space. Hence, in order that a topological linear space be metrizable, it is *necessary* that it be a T_1-space and that there exist a countable base at 0. These necessary conditions are also *sufficient*, not merely to guarantee that the space is metrizable, but that it be metrizable with an invariant metric. We state this formally:

Theorem 3.9–B. *Let X be a T_1-topological linear space in which there exists a countable base at* 0. *Then X is metrizable with an invariant metric.*

We omit the rather lengthy proof of this theorem. See Kelley, **1**, pages 185–186 and p. 210.

In a metric linear space the concept of boundedness, as defined in § 3.3 for a topological linear space, does not need to coincide with the concept of boundedness for a metric space, as defined in § 2.4. A set may be metrically bounded but not bounded in the sense of § 3.3. This is easily shown to be possible in the space (s) of Example 1. In fact, the space as a whole is metrically bounded, but not bounded as a topological linear space.

PROBLEMS

1. A topological linear T_1-space is metrizable if there exists a bounded neighborhood of 0. The converse is not true. A space is called *locally bounded* if there exists in it a bounded neighborhood of 0.

2. Let X be a linear space, and suppose M is a subspace of X^f. Then the topology $\mathscr{T}(X, M)$ defined in § 3.81 is metrizable if and only if M is total and possesses an at most countable Hamel basis. The topology is normable if and only if M is total and finite dimensional.

3. Consider $L^p = L(a, b)$, where (a, b) is any interval of the real axis and $0 < p < 1$. For $x \in L^p$ let

$$|x| = \int_a^b |x(s)|^p \, ds.$$

Then L^p becomes a complete locally bounded metric linear space according to the prescription of Theorem 3.9–A. For the inequality $|x + y| \leqslant |x| + |y|$ see Hardy, Littlewood and Pólya **1**, Theorem 199, page 147. The space is not locally convex; in fact, the only nonempty open convex set in L^p is the whole space (see problem 4, § 3.6).

4. Consider H^p, where $0 < p < 1$ (see Example 8, § 3.11). For $f \in H^p$ define

$$|f| = \sup_{0 \leqslant r < 1} \int_0^{2\pi} |f(re^{i\theta})|^p \, d\theta = \int_0^{2\pi} |f(e^{i\theta})|^p \, d\theta.$$

Then H^p may be regarded as a closed subspace of $L^p(0, 2\pi)$. In contrast to the situation for L^p, H^p possesses many continuous nonzero linear functionals. In fact, for a fixed z with $|z| < 1$, $f(z)$ is a continuous linear functional of f. Also, the Taylor coefficients in the power series expansion of f are such functionals. For discussion of H^p when $0 < p < 1$, see Walters, **1**, **2**. A. E. Livingston, **1**, has shown that the space is not locally convex and hence certainly not normable.

5. If x' is a continuous linear functional defined on the space (s), there exists a finite set $\alpha_1, \cdots, \alpha_N$ such that $x'(x) = \sum_{n=1}^{N} \alpha_n \xi_n$ for each $x \in (s)$.

SUGGESTED READING

For normed linear spaces

Banach, **1**, especially Chapter 4.
Dunford and Schwartz, **1**, Chapters 1, 4, 5.
Fréchet, **1**.
Ljusternik and Sobolew, **1**, Chapter 2.
Zaanen, **1**, Chapter 6.

For inner-product spaces

Achieser and Glasmann, **1**, Chapter 1.
Bourbaki, **4**, especially Chapter 5 of No. 1229.
Cooke, **1**, Chapter 1.
Halmos, **3**, Chapter 1.
Jacobson, **1**, Chapter 6.
Julia, **1**, Chapters 1–3.
Nagy, **1**, Chapter 1.
Schmeidler, **1**, Chapter 1.
Stone, **1**, Chapters 1–3.
von Neumann, **1**, Chapters 1–2.
Zaanen, **1**, Chapter 6.

For topological linear spaces

Bourbaki, **4**.
Dieudonné, **1**, **2**.
Hyers, **1**.
Mackey, **1**, **2**.
von Neumann, **2**.
Wehausen, **1**.

4

GENERAL THEOREMS ON LINEAR OPERATORS

4.0 Introduction

In this chapter the emphasis is principally on the study of continuous linear operators in normed linear spaces. Completeness of the spaces is assumed only where it seems to be essential for decisive results. Some important results, especially those concerning closed operators in § 4.2, are valid in complete metric linear spaces, where the metric need not be that defined by a norm. The middle portion of the chapter (§§ 4.5–4.71) is concerned with conjugate spaces and conjugate operators and presents the more intricate counterpart, for Banach spaces, of the purely algebraic study made in §§ 1.8–1.91 of the relations between an operator and its transpose. The later part of the chapter, beginning with § 4.8, is devoted to inner-product spaces, the special aim being to present some of the fundamental facts about continuous linear functionals and the adjoint of a bounded linear operator in Hilbert space. The theory of adjoint operators closely resembles the theory of conjugate operators but has certain special features because of the fact that the normed conjugate space of a complete inner-product space is congruent in a special way to the space itself.

There are many far-reaching results in this chapter. The massive

Theorem 4.2–G is the fount from which spring many of the most important results in the theory of linear spaces. Theorems 4.2–H and 4.2–I, both very important, follow directly from 4.2–G. Theorem 4.2–I is the often used "closed-graph theorem." From it we deduce the highly important and useful theorems which express, in one form or another, the principle of uniform boundedness. Our most general formulations of this principle are given in Theorems 4.4–B and 4.4–E.

The normed-linear-space form of the Hahn-Banach theorem is given in 4.3–A. This theorem is an indispensable tool in the proofs of many subsequent theorems. As an illustration of its more direct use in analysis we show, at the end of § 4.3, how the Hahn-Banach theorem can be used to give an existence proof in connection with the Dirichlet problem.

Weak topologies are of significant importance in the theory of normed linear spaces. We mention here especially Theorems 4.61–A, 4.61–C, and 4.62–A.

A large block of material in this chapter deals with a bounded linear operator and its conjugate. There are many theorems, and it is hard to single out any selected few of them as being the most important. Those in § 4.7 are more profound than those in § 4.6. The pooling of all the results about an operator and its conjugate, using theorems from §§ 4.2, 4.6, and 4.7, leads to the "state diagram" in § 4.71, which is a complete tabulation of all the relations of a certain kind which can exist between an operator and its conjugate.

In the last part of the chapter the central theorems are the representation theorem for continuous linear functionals on a Hilbert space (4.81–C), and the closely associated Theorem 4.82–A, on orthogonal complements (often called "the projection theorem").

4.1 Spaces of Linear Operators

Linear operators were defined in § 1.3. If X and Y are linear spaces (with the same scalar field), the set of all linear operators on X into Y is a linear space if we define addition of operators and multiplication of operators by scalars in the natural way, namely

$$(A + B)x = Ax + Bx, \qquad (\alpha A)x = \alpha(Ax).$$

The operator which maps every $x \in X$ into the zero element of Y is the zero of the linear space of operators; without serious danger of confusion with other zeros we denote this zero operator by 0.

If X and Y are topological linear spaces, those linear operators on X into Y which are continuous on X form a subspace of the space of *all* linear operators on X into Y. We shall denote the linear space of all *continuous* linear operators on X into Y by $[X, Y]$.

If $Y = X$, the space of all linear operators on X into Y is not merely a linear space; it is an *algebra*, according to the following definition of that term.

Definition. A set X is called an algebra if it is a linear space in which to each ordered pair x, y corresponds an element of X denoted by xy and called the product "x times y," subject to the following axioms:

1. $(xy)z = x(yz)$.
2. $x(y + z) = xy + xz$.
3. $(x + y)z = xz + yz$.
4. $(\alpha x)(\beta y) = (\alpha\beta)(xy)$.

Here it is understood that x, y, z are arbitrary elements of X and α, β are arbitrary scalars. The algebra is called real or complex according as the scalar field is real or complex.

If the algebra X contains an element e such that $ex = xe = x$ for every x, e is called the *unit* of the algebra; it is necessarily unique if it exists.

An algebra is called commutative if $xy = yx$ for every pair x, y. An algebra need not be commutative.

If X_0 is a subset of an algebra X and if $x + y$, αx and xy are in X_0 whenever x, y are any elements of X_0 and α is any scalar, then X_0 is called a *subalgebra* of X. Observe that X_0 is also a linear manifold, or subspace, in X.

If A and B are linear operators on X into X, where X is any linear space, we define AB as the linear operator such that $(AB)x = A(Bx)$ for each $x \in X$. With this definition it is clear that the set of *all* linear operators on X into X is an algebra. Moreover, this algebra has a unit, namely, the operator I defined by $Ix = x$ for every x. The *continuous* linear operators on X into X also form an algebra with this same unit I. Instead of the notation $[X, X]$ we shall use the simpler notation $[X]$ for this algebra of all continuous linear operators on X into X.

If X and Y are normed linear spaces, a linear operator A on X into Y is continuous on X (i.e., $A \in [X, Y]$) if $\sup_{\|x\| \leqslant 1} \|Ax\| < \infty$. For such an A we define

$$\|A\| = \sup_{\|x\| \leqslant 1} \|Ax\|. \tag{4.1–1}$$

These matters were discussed in § 3.1; see especially Theorem 3.1–A and formulas (3.1–1)–(3.1–3); the latter formulas give useful alternative formulas for $\|A\|$.

Theorem 4.1–A. *If X and Y are normed linear spaces, $[X, Y]$ is a normed linear space with the norm defined by* (4.1–1). *If Y is complete so is $[X, Y]$.*

PROOF. We leave to the reader the simple verification that $\|A\|$ is a norm on $[X, Y]$. We shall prove the assertion about completeness. Let $\{A_n\}$ be a Cauchy sequence in $[X, Y]$. For given $\epsilon > 0$ let N be such that $\|A_m - A_n\| < \epsilon$ if $N \leqslant m$ and $N \leqslant n$. Then also, for each x, $\|A_m x - A_n x\| \leqslant \|A_m - A_n\| \, \|x\| < \epsilon\|x\|$. It is thus clear that $\{A_n x\}$ is a Cauchy sequence in Y. Since Y is complete, $A_n x$ has a limit which depends on x; we define $Ax = \lim\limits_{n\to\infty} A_n x$. It is readily seen that A is a linear operator on X into Y. Now $\{\|A_n\|\}$ is a bounded sequence (since every Cauchy sequence in a normed linear space is bounded), and we have $\|A_n\| \leqslant M$ for all n, where M is some constant. Then $\|A_n x\| \leqslant M\|x\|$ for every x; it follows that $\|Ax\| \leqslant M\|x\|$. Therefore A is a member of $[X, Y]$, by Theorem 3.1–A. If we now return to the inequality $\|A_m x - A_n x\| < \epsilon\|x\|$ and let $n \to \infty$, we see that $\|A_m x - Ax\| \leqslant \epsilon\|x\|$ for every x if $N \leqslant m$. This means that $\|A_m - A\| \leqslant \epsilon$ if $N \leqslant m$. But then $A_m \to A$ in the topology defined on $[X, Y]$ by the norm. Consequently $[X, Y]$ is complete.

Suppose that A_n and A are in $[X, Y]$ and that $\|A_n - A\| \to 0$. This is equivalent to the statement that $\|A_n x - Ax\| \to 0$ uniformly for all $x \in X$ such that $\|x\| \leqslant 1$. For this reason the topology for $[X, Y]$ defined by the norm (4.1–1) is sometimes called the *uniform topology* for $[X, Y]$.

For the case $Y = X$, $[X, Y] = [X]$, we have the following theorem:

Theorem 4.1–B. *Suppose X is a normed linear space. Then*

$$\|AB\| \leqslant \|A\| \, \|B\| \tag{4.1–2}$$

for each pair A, B in $[X]$. Also

$$\|I\| = 1. \tag{4.1–3}$$

PROOF. For each x, $\|(AB)x\| = \|A(Bx)\| \leqslant \|A\| \, \|Bx\| \leqslant \|A\| \, \|B\| \, \|x\|$, and (4.1–2) is a consequence. The truth of (4.1–3) is immediate from the definition of the norm.

From (4.1–2) it follows that AB is a continuous function of the pair A, B.

When X and Y are normed linear spaces, it is a common practice to call the elements of $[X, Y]$ *bounded linear operators*. The use of the adjective "bounded" as an equivalent for "continuous" is explained by the remarks following formula (3.1–3).

The next two theorems furnish very important information about the existence and nature of inverse operators under certain conditions. Formula (4.1–4) is closely related to an old result in the theory of integral equations, generally known as the Neumann expansion (named after C. Neumann, cf. Hellinger and Toeplitz, **1**, page 1347), though the method goes back as far as Liouville.

Theorem 4.1–C. *Let X be a Banach space, and suppose $A \in [X]$ and $\|A\| < 1$. Then the range of $I - A$ is X, $(I - A)^{-1}$ exists and belongs to $[X]$, and*

$$(I - A)^{-1} = I + A + A^2 + \cdots + A^n + \cdots, \tag{4.1–4}$$

the series converging in the uniform topology of the space $[X]$. Furthermore,

$$\|(I - A)^{-1}\| \leqslant \frac{1}{1 - \|A\|}. \tag{4.1–5}$$

PROOF. Since $\|A\| < 1$, the series $\sum_0^\infty \|A\|^n$ converges. But $\|A^n\| \leqslant \|A\|^n$, by (4.1–2), and hence the series $\sum_0^\infty A^n$ converges in $[X]$, owing to the completeness of the latter space (see Theorem 4.1–A). If we let $B = \sum_0^\infty A^n$, we see that

$$AB = BA = \sum_0^\infty A^{n+1},$$

and hence

$$(I - A)B = B(I - A) = I.$$

This implies that $(I - A)^{-1}$ exists and is equal to B. The inequality (4.1–5) is now evident, and so the proof is complete.

Theorem 4.1–D. *Let X be a Banach space. Let R be the set of elements $A \in [X]$ such that $\mathscr{R}(A) = X$ and A^{-1} exists and is continuous. If $A \in R$, $B \in [X]$, and $\|A - B\| < 1/\|A^{-1}\|$, then $B \in R$, and*

$$\|B^{-1}\| \leqslant \frac{\|A^{-1}\|}{1 - \|A^{-1}\|\,\|A - B\|}, \tag{4.1–6}$$

$$\|B^{-1} - A^{-1}\| \leqslant \frac{\|A^{-1}\|^2\|A - B\|}{1 - \|A^{-1}\|\,\|A - B\|}. \tag{4.1–7}$$

Thus, in particular, R is an open set in $[X]$ and A^{-1} is a continuous function of A.

PROOF. We can write

$$B = A - (A - B) = A[I - A^{-1}(A - B)].$$

Since $\|A^{-1}(A - B)\| \leqslant \|A^{-1}\| \, \|A - B\| < 1$, it follows by Theorem 4.1–C that $[I - A^{-1}(A - B)]^{-1}$ exists, is in $[X]$, and is given by

$$[I - A^{-1}(A - B)]^{-1} = \sum_{n=0}^{\infty} \{A^{-1}(A - B)\}^n. \tag{4.1–8}$$

It follows readily that B^{-1} exists, belongs to $[X]$, and is given by

$$B^{-1} = [I - A^{-1}(A - B)]^{-1}A^{-1}.$$

The inequalities (4.1–6) and (4.1–7) follow readily from the series (4.1–8), and we omit the details. It may be remarked that

$$\|B^{-1} - A^{-1}\| \leqslant \frac{\|A^{-1}\| \, \|I - A^{-1}B\|}{1 - \|I - A^{-1}B\|}.$$

This will in general be a sharper inequality than (4.1–7), since $\|I - A^{-1}B\| = \|A^{-1}(A - B)\| \leqslant \|A^{-1}\| \, \|A - B\|$. Likewise (4.1–6) can be sharpened.

PROBLEM

Suppose that X, Y, and Z are normed linear spaces, that X is a subspace of Z dense in Z, and that Y is complete. Show that $[X, Y]$ is congruent to $[Z, Y]$ by mapping $T \in [Z, Y]$ into $A \in [X, Y]$, where A is the restriction of T to X.

4.11 Integral Equations of the Second Kind. The Neumann Expansion

We start out by referring back to Example 2 in § 1.5, where integral equations of Fredholm and Volterra type were defined. Such equations can be considered in various function spaces; at first we shall deal with the space of continuous functions. Let X be the Banach space $C[a, b]$, and let K be the element of $[X]$ defined by $Kx = y$, where

$$y(s) = \int_a^b k(s, t)x(t)\, dt, \tag{4.11–1}$$

and the function $k(s, t)$ (called the kernel) is subjected to certain restrictions. There are various sets of restrictions which are adequate for the purposes of the developments which are to follow. The simplest restrictions are as follows:

a. $k(s, t)$ is continuous in both variables when $a \leqslant s \leqslant b$ and $a \leqslant t \leqslant b$.

b. If M is the maximum of $|k(s, t)|$ on the square $[a, b] \times [a, b]$, then $M(b - a) < 1$.

These restrictions may be made less severe, but at the cost of some complication in statements and proofs. Our main purpose is to give an exposition of the formal aspects of part of the theory of integral equations of the second kind, and the conditions *a*, *b* will serve conveniently for the rigorous justification of the formal procedures.

We are concerned with the equation

$$y(s) = x(s) - \int_a^b k(s, t)x(t)\, dt;$$

as an equation in the space X it can be written

$$y = x - Kx. \tag{4.11–2}$$

According to Theorem 4.1–C, this equation has a unique solution x for each given y if $\|K\| < 1$. (We do not assert that $\|K\| < 1$ is a *necessary* condition; only that it is sufficient.) The solution is

$$x = y + Ky + K^2y + \cdots + K^ny + \cdots. \tag{4.11–3}$$

Now, from (4.11–1) and the definition of M in condition *b*, we see that

$$\|K\| \leqslant M(b - a). \tag{4.11–4}$$

Condition *b* then assures us that $\|K\| < 1$.

We are going to show that each of the operators $K^2, K^3, \cdots$ is of the same sort as K, with a suitable kernel. We shall define

$$\begin{cases} k_1(s, t) = k(s, t) \\ k_n(s, t) = \displaystyle\int_a^b k_1(s, u)k_{n-1}(u, t)\, du, \quad n > 1. \end{cases} \tag{4.11–5}$$

By definition, $z = K^2y$ means $z = K(Ky)$, or

$$z(s) = \int_a^b k(s, t)\, dt \int_a^b k(t, u)y(u)\, du.$$

We invert the order of integration, obtaining

$$z(s) = \int_a^b y(u)\, du \int_a^b k(s, t)k(t, u)\, dt,$$

$$z(s) = \int_a^b k_2(s, u)y(u)\, du.$$

Thus K^2 is defined in terms of the kernel k_2 just as K is defined in terms of k_1. It is easily proved by induction that $z = K^n y$ is expressed by

$$z(s) = \int_a^b k_n(s, t)y(t)\,dt, \tag{4.11–6}$$

so that k_n is the kernel used to determine K^n. For $n > 1$, k_n is called the nth iterated kernel.

The series (4.11–3) now takes the form

$$x(s) = y(s) + \sum_{n=1}^{\infty} \int_a^b k_n(s, t)y(t)\,dt. \tag{4.11–7}$$

The convergence of (4.11–3) as a series of elements in the Banach space $C[a, b]$ means that (4.11–7) is uniformly convergent on $[a, b]$. Now consider the function h defined by

$$h(s, t) = \sum_{n=1}^{\infty} k_n(s, t). \tag{4.11–8}$$

It is easily proved by induction from (4.11–5) that

$$|k_n(s, t)| \leqslant M^n(b - a)^{n-1}, \qquad n > 1.$$

Since we have assumed $M(b - a) < 1$, it follows that the series (4.11–8) converges uniformly; thus h is a continuous function of s and t. Moreover, we can exchange the order of summation and integration in (4.11–7) and obtain

$$x(s) = y(s) + \int_a^b h(s, t)y(t)\,dt. \tag{4.11–9}$$

We shall define an operator H with h as kernel, so that $z = Hy$ means

$$z(s) = \int_a^b h(s, t)y(t)\,dt.$$

Then (4.11–9) becomes

$$x = y + Hy.$$

We thus arrive at the conclusion that, when conditions a and b are satisfied, there is an operator H determined by a kernel h in such a way that

$$(I - K)^{-1} = I + H. \tag{4.11–10}$$

The kernel h is sometimes called the *reciprocal* kernel corresponding to k. It has also been called the *resolvent* kernel, but in modern operator theory the word resolvent is used in a slightly different way.

An operator K defined by a kernel k as in (4.11–1) is called a Fredholm-type operator. The name goes back to the Swede, Fredholm, who developed a comprehensive theory for integral equations of the second kind at the beginning of the twentieth century.

The theory of Volterra integral equations may be developed independently, or it may be regarded as a special case, with important special features, of the theory of Fredholm integral equations. A Volterra-type operator K is defined by $y = Kx$, where

$$y(s) = \int_a^s k(s, t)x(t)\, dt.$$

We shall assume that the kernel $k(s, t)$ is continuous in both variables in the triangular region defined by the inequalities $a \leqslant t \leqslant s \leqslant b$. If we define k to have the value 0 in the part of the square $[a, b] \times [a, b]$ where $t > s$, we can regard K as a Fredholm-type operator with this special feature of the kernel k. Since $k(s, s)$ need not be 0, there may be discontinuities of the kernel at points along the diagonal $t = s$ of the square. These discontinuities have no adverse effects on the earlier developments in this section, however.

It may be verified by induction from (4.11–5) that, when k is a Volterra kernel, the iterated kernels k_n are also Volterra kernels; that is, $k_n(s, t) = 0$ if $t > s$. The iterative formula for k_n becomes

$$(4.11\text{–}11) \qquad k_n(s, t) = \int_t^s k_1(s, u)k_{n-1}(u, t)\, du, \qquad n > 1, \qquad t \leqslant s.$$

From this it can readily be proved by induction that

$$(4.11\text{–}12) \qquad |k_n(s, t)| \leqslant M^n \frac{(s - t)^{n-1}}{(n - 1)!}$$

where M is the maximum of $|k(s, t)|$. Hence certainly

$$|k_n(s, t)| \leqslant M^n \frac{(b - a)^{n-1}}{(n - 1)!}.$$

This inequality shows that the series (4.11–8) converges uniformly. The function $h(s, t)$ is also a Volterra kernel; it is continuous in the triangle $a \leqslant t \leqslant s \leqslant b$, and $h(s, t) = 0$ if $t > s$.

From (4.11–12) and the formula for K^n in terms of k_n it follows that

$$(4.11\text{–}13) \qquad \|K^n\| \leqslant M^n \frac{(b - a)^n}{n!}.$$

Consequently the series $\sum_{n=0}^{\infty} K^n$ converges in the uniform topology of the space $[X]$. A perusal of the proof of Theorem 4.1–C shows that the assertion made in connection with (4.1–4) remains valid under any conditions which insure the convergence of the series in (4.1–4). As a consequence, it follows that the equation $y = x - Kx$ has a unique solution for each y, the solution being

$$x = \sum_{n=0}^{\infty} K^n y.$$

Furthermore, this solution may be expressed in the form

$$x(s) = y(s) + \int_a^s h(s, t) y(t)\, dt.$$

The very important fact about Volterra integral equations of the second kind is, then, that such an equation is always uniquely solvable. The special nature of the Volterra-type operator K insures that the Neumann expansion is always convergent, regardless of the magnitude of the kernel k. This is not true in general in the case of a Fredholm-type operator.

4.12 $\mathscr{L}^2$ Kernels

In this section we consider the space $L^2(a, b)$ (hereafter in this section written simply as L^2). Suppose $k(s, t)$ belongs to the class $\mathscr{L}^2(E)$, where E is the square $[a, b] \times [a, b]$. Then we call k an $\mathscr{L}^2$ kernel. Such a kernel defines a bounded operator K on L^2 in the following way: If $x \in L^2$ let $y = Kx$, where this means

$$y(s) =^{0} \int_a^b k(s, t) x(t)\, dt.$$

Since

$$|y(s)|^2 \leqslant^{0} \int_a^b |k(s, t)|^2\, dt \int_a^b |x(t)|^2\, dt,$$

we see that

$$\int_a^b |y(s)|^2\, ds \leqslant \int_a^b \int_a^b |k(s, t)|^2\, ds\, dt \int_a^b |x(t)|^2\, dt.$$

This shows that

$$\|y\| = \|Kx\| \leqslant \left(\int_a^b \int_a^b |k(s, t)|^2\, ds\, dt \right)^{\frac{1}{2}} \|x\|,$$

and hence that

$$\|K\| \leqslant \left(\int_a^b \int_a^b |k(s, t)|^2\, ds\, dt \right)^{\frac{1}{2}}. \tag{4.12–1}$$

As a member of $\mathscr{L}^2(E)$, $k(s, t)$ determines an element of the Hilbert space $L^2(E)$. We denote this element by k and its norm in $L^2(E)$ by $\|k\|$:

$$\|k\| = \left(\int_a^b \int_a^b |k(s, t)|^2 \, ds \, dt\right)^{\frac{1}{2}}. \tag{4.12–2}$$

From (4.12–1) we see that $\|K\| \leqslant \|k\|$.

The equation $y = x - Kx$ in L^2 corresponds to a Fredholm integral equation

$$y(s) =^0 x(s) - \int_a^b k(s, t)x(t) \, dt$$

in $\mathscr{L}^2$. By Theorem 4.1–C we know that $y = x - Kx$ has a unique solution x in L^2 for each choice of y in L^2, provided that $\|K\| < 1$. This will certainly be the case if $\|k\| < 1$.

In the case of $\mathscr{L}^2$ kernels we can define the iterated kernels k_n and the reciprocal kernel h in a manner formally the same as we did in § 4.11 with continuous kernels. The details are somewhat different, however. The iterated kernels are defined essentially as before: $k_1(s, t) = k(s, t)$ and

$$k_n(s, t) =^0 \int_a^b k_1(s, u)k_{n-1}(u, t) \, du,$$

where $=^0$ refers to the exception of a set of two-dimensional measure zero. It is not difficult to show by induction that we also have

$$k_n(s, t) =^0 \int_a^b k_{n-1}(s, u)k(u, t) \, du. \tag{4.12–3}$$

Also, by induction and the Schwarz inequality we find that k_n is an $\mathscr{L}^2$ kernel, with

$$\int_a^b \int_a^b |k_n(s, t)|^2 \, ds \, dt \leqslant \left(\int_a^b \int_a^b |k(s, t)|^2 \, ds \, dt\right)^n,$$

or

$$\|k_n\| \leqslant \|k\|^n.$$

It is a more delicate matter to estimate the numerical magnitude of $k_n(s, t)$. Let us write

$$u(s) =^0 \left(\int_a^b |k(s, t)|^2 \, dt\right)^{\frac{1}{2}}$$

$$v(t) =^0 \left(\int_a^b |k(s, t)|^2 \, ds\right)^{\frac{1}{2}}.$$

Now, for each t except on a set of measure 0, $k(s, t)$ as a function of s

determines an element k_t of L^2. Likewise $k_n(s, t)$ determines an element $k_t^{(n)}$. Formula (4.12–3) shows that

$$k_t^{(n)} = K^{n-1}k_t = Kz,$$

where $z = K^{n-2}k_t$. Thus

$$k_n(s, t) =^0 \int_a^b k(s, u)z(u)\, du,$$

$$|k_n(s, t)| \leqslant^0 \left(\int_a^b |k(s, u)|^2\, du\right)^{\frac{1}{2}} \left(\int_a^b |z(u)|^2\, du\right)^{\frac{1}{2}},$$

$$|k_n(s, t)| \leqslant^0 u(s)\, \|z\|.$$

But

$$\|z\| = \|K^{n-2}k_t\| \leqslant \|K\|^{n-2}\, \|k_t\| = \|K\|^{n-2}v(t),$$

and so

$$(4.12\text{–}4) \qquad |k_n(s, t)| \leqslant^0 \|K\|^{n-2}u(s)v(t), \qquad n \geqslant 2.$$

This inequality shows that the series (4.11–8) converges and defines a function $h(s, t)$ almost everywhere on $E = [a, b] \times [a, b]$, provided that $\|K\| < 1$. The inequality (4.12–4) also enables us to conclude that h is an $\mathscr{L}^2$ kernel, that the series (4.11–7) converges almost everywhere, and that we can reverse the order of summation and integration to obtain

$$x(s) =^0 y(s) + \int_a^b h(s, t)y(t)\, dt$$

as the solution of the integral equation in $\mathscr{L}^2$ (all this when $\|K\| < 1$).

Observe that in this case the Banach space theory tells us merely that

$$x = y + Ky + K^2y + \cdots + K^ny + \cdots,$$

the series being convergent in L^2. The additional information that

$$x(s) = y(s) + \int_a^b k(s, t)y(t)\, dt + \cdots + \int_a^b k_n(s, t)y(t)\, dt + \cdots$$

is convergent almost everywhere, with its terms dominated as in (4.12–4), is more than was obtained directly from Theorem 4.1–C.

4.13 Differential Equations and Integral Equations

In order to indicate some of the reasons why integral equations are of significant interest, we shall show how certain types of questions concerning differential equations can be recast as questions about integral equations. For simplicity we confine attention to second-order linear ordinary differential equations. The discussion could be extended to cover ordinary equations of higher order.

Consider the equation

$$y''(s) + a_1(s)y'(s) + a_2(s)y(s) = x(s), \tag{4.13–1}$$

where a_1 and a_2 are fixed continuous functions of s, defined when $a \leqslant s \leqslant b$. Let $X = C[a, b]$. The *initial-value problem* for (4.13–1) is the problem of finding $y \in X$, with y' and y'' also in X, so that (4.13–1) is satisfied and so that $y(a)$ and $y'(a)$ have preassigned values, say $y(a) = \alpha_0$, $y'(a) = \alpha_1$. The function x is to be an arbitrarily assigned member of X. The fact that this initial-value problem always has a solution (and a unique one), no matter how α_0, α_1, and x are chosen, is a consequence of standard existence theorems in the theory of differential equations. But we can also demonstrate the fact by using what we have learned in § 4.11 about Volterra integral equations.

We write $y''(s) = z(s)$, assuming for the moment that y is a solution of the initial value problem. Then

$$\int_a^u z(t)\,dt = y'(u) - \alpha_1,$$

$$\int_a^s du \int_a^u z(t)\,dt = y(s) - \alpha_0 - \alpha_1(s - a). \tag{4.13–2}$$

The iterated integral in this last formula can be expressed in the form

$$\int_a^s z(t)\,dt \int_t^s du = \int_a^s (s - t)z(t)\,dt.$$

On replacing y, y', and y'' in (4.13–1) by their expression in terms of z, we obtain

$$z(s) + a_1(s)\left\{\alpha_1 + \int_a^s z(t)\,dt\right\} + a_2(s)\left\{\alpha_0 + \alpha_1(s - a) + \int_a^s (s - t)z(t)\,dt\right\} = x(s). \tag{4.13–3}$$

This equation is satisfied by z if y is a solution of the initial-value problem. Conversely, if z satisfies (4.13–3) and y is defined by (4.13–2), y is a solution of the initial-value problem.

If we define

$$k(s, t) = -a_1(s) - a_2(s)(s - t),$$

$$w(s) = x(s) - \alpha_0 a_2(s) - \alpha_1\{a_1(s) + a_2(s)(s - a)\},$$

the equation (4.13–3) can be written in the form

$$z(s) - \int_a^s k(s, t)z(t)\,dt = w(s).$$

This is a Volterra integral equation of the second kind. As we know from § 4.11, it has a unique solution $z \in X$ for each $w \in X$. *Therefore, the initial value problem for* (4.13–1) *always has a unique solution.*

There is another important class of problems related to the differential equation (4.13–1), namely the two-point problems. For a *two-point problem* we ask for a solution of the differential equation which satisfies a certain specified condition at $s = a$ and another condition at $s = b$. A common form of condition is one in which we specify the value at an end point of some predetermined linear combination of y and y'. We shall illustrate by discussing the problem with the conditions $y(a) = \alpha$, $y(b) = \beta$. We refer back to Example 3 of § 1.3 for certain things which we need. There it was shown (we change the notation slightly) that if $x \in C[a, b]$ and if

$$c(s, t) = \begin{cases} \dfrac{(b - s)(t - a)}{b - a} & \text{when } a \leqslant t \leqslant s \\[2ex] \dfrac{(s - a)(b - t)}{b - a} & \text{when } s \leqslant t \leqslant b, \end{cases}$$

then the function

$$z(s) = -\int_a^b c(s, t)x(t)\,dt$$

satisfies the conditions

$$z''(s) = x(s), \qquad z(a) = z(b) = 0.$$

Consequently the function

$$y(s) = \alpha\frac{b - s}{b - a} + \beta\frac{s - a}{b - a} - \int_a^b c(s, t)x(t)\,dt \tag{4.13–4}$$

satisfies the conditions

$$y''(s) = x(s), \qquad y(a) = \alpha, \qquad y(b) = \beta.$$

In (4.13–4) let us replace $x(t)$ by

$$x(t) - a_1(t)y'(t) - a_2(t)y(t).$$

In this way we see that y will be a solution of our two-point problem if and only if it has a continuous derivative and satisfies the equation

$$\begin{aligned} y(s) = \alpha\frac{b - s}{b - a} &+ \beta\frac{s - a}{b - a} \\ &- \int_a^b c(s, t)\{x(t) - a_1(t)y'(t) - a_2(t)y(t)\}\,dt. \end{aligned} \tag{4.13–5}$$

The discussion of this equation is much simplified if we assume that $a_1(t) \equiv 0$. For this special case let us write

$$k(s, t) = c(s, t)a_2(t),$$

$$w(s) = \alpha \frac{b - s}{b - a} + \beta \frac{s - a}{b - a} - \int_a^b c(s, t)x(t)\, dt.$$

Then (4.13–5) becomes

$$y(s) - \int_a^b k(s, t)y(t)\, dt = w(s).$$

This is a Fredholm integral equation of the second kind. It will certainly have a unique solution $y \in C[a, b]$ if $\|K\| < 1$, where K is the Fredholm-type operator with kernel k. (The condition $\|K\| < 1$ is sufficient; it may not be necessary.) We know from § 4.11 that $\|K\| \leqslant M(b - a)$, where M is the maximum value of $|k(s, t)|$. Now, the maximum value of $(b-a)\,|c(s, t)|$ is readily found to be $(b-a)^2/4$. Let A be the maximum value of $|a_2(t)|$. Then certainly $\|K\| < 1$ if $(b - a)^2A < 4$. We summarize the conclusions:

The two-point problem for the equation

$$y''(s) + a_2(s)y(s) = x(s)$$

with the end conditions $y(a) = \alpha$, $y(b) = \beta$, *is always uniquely solvable, with arbitrarily assigned* α, β, x, *provided that*

$$(b - a)^2 \max |a_2(s)| < 4.$$

(For what may happen if this inequality is not satisfied, see problem 14 in § 5.5.)

We return to the case when $a_1(t) \not\equiv 0$. One way to handle this is to make a change of variable. If $a_1'(s)$ is continuous, we can define

$$q(s) = e^{-\frac{1}{2}\int_a^s a_1(t)\, dt}$$

and set $y = uq$. Then, with u as the new unknown, we have the equation

$$u''(s) + \{a_2(s) - \tfrac{1}{2}a_1'(s) - \tfrac{1}{4}[a_2(s)]^2\}u(s) = \frac{x(s)}{q(s)}.$$

The new end conditions are

$$u(a) = \frac{\alpha}{q(a)} = \alpha, \qquad u(b) = \frac{\beta}{q(b)}.$$

Since the $u'(s)$ term does not appear in the differential equation, the previous method may be used to recast the problem as an integral equation problem.

If a_1 cannot be eliminated, so that the y' term is actually present in (4.13–5), we have what is called an *integro-differential* equation. It is a more complicated matter to discuss the possibility of solution of such an equation, and we shall leave the matter at this point.

In some later sections (§ 6.11 and § 6.41) we shall discuss another phase of the relation between differential and integral equations, with applications to eigenvalue problems of differential equations of Sturm-Liouville type.

4.2 Closed Linear Operators

For the purposes of applications to analysis it is essential to consider some linear operators which are not continuous. However, many of the most important discontinuous linear operators have a property which in some respects compensates for the absence of the property of continuity. This property is most naturally described in terms of the concept of the graph of a function.

Definition. Let X and Y be topological spaces. Let f be a function with domain $\mathscr{D}$ and range $\mathscr{R}$ such that $\mathscr{D} \subset X$ and $\mathscr{R} \subset Y$. The set $\{(x, f(x)) : x \in \mathscr{D}\}$ in the product space $X \times Y$ is called the *graph* of f. The function is called *closed* in case the graph is a closed set in $X \times Y$ (with the topology of the product space as defined in § 2.5).

Observe that the graph of f is the set of all ordered pairs $(x, f(x))$, where $x \in \mathscr{D}$.

In case X and Y are metric spaces, the topology of $X \times Y$ can also be defined by a metric. From Theorem 2.4–A we obtain the following theorem:

Theorem 4.2–A. *If X and Y are metric spaces a function f with domain $\mathscr{D} \subset X$ and range $\mathscr{R} \subset Y$ is closed if and only if the situation $x_n \in \mathscr{D}$, $x_n \to x$, $f(x_n) \to y$ implies that $x \in \mathscr{D}$ and $f(x) = y$.*

For many purposes the criterion for a closed function which we find it convenient to use is that given in Theorem 4.2–A.

Example 1. *A discontinuous but closed operator*: Let $X = C[0, 1]$ and let $X = Y$. Let $\mathscr{D}$ be the set of $x \in X$ such that the derivative $x'(t)$ is defined and continuous on $[0, 1]$; let T be the operator with domain $\mathscr{D}$ defined by $Tx = x'$. It is evident that T is linear. If $x_n(t) = t^n$, then $\|x_n\| = 1$, $x_n'(t) = nt^{n-1}$, and so $\|Tx_n\| = \|x_n'\| = n$ $(n = 1, 2, \cdots)$. Thus T is not continuous, because $\sup\limits_{\|x\|=1} \|Tx\| = \infty$. But T is closed. In fact, suppose $x_n \in \mathscr{D}$, $x_n \to x$, $Tx_n \to y$. Then $x_n'(t)$ converges uniformly to

$y(t)$, and y is continuous. It follows by a standard convergence theorem that $x(t)$ is differentiable, with derivative equal to $y(t)$. Therefore $x \in \mathscr{D}$, $Tx = y$, and T is closed.

One of the important fields of application of functional analysis is in the theory of differential equations. Operators defined by means of ordinary or partial differentiation are frequently discontinuous, but in the applications it is often possible to arrange matters so that the differential operators to be dealt with are closed.

Example 2. Let $X = L^2(0, 1)$. Let T be an operator with domain and range in X, defined as follows: $\mathscr{D}(T)$ is the class of those $x \in X$ for which a representative function in $\mathscr{L}^2(0, 1)$ can be found such that $x(s)$ is absolutely continuous on $[0, 1]$, $x(0) = x(1)$, and $x'(s)$ is in $\mathscr{L}^2(0, 1)$. This means that $x(s)$ is of the form

$$x(s) = \alpha + \int_0^s u(t)\,dt,$$

where $u(t) \in \mathscr{L}^2(0, 1)$, α is a scalar, and

$$\int_0^1 u(t)\,dt = 0.$$

We then define $Tx = x'$.

We shall demonstrate that T is closed. Assume that $x_n \in \mathscr{D}$, $x, y \in X$, $x_n \to x$, $Tx_n \to y$. We can write

$$x_n(s) = x_n(0) + \int_0^s x_n'(t)\,dt. \tag{4.2–1}$$

Now

$$\left|\int_0^s x_n'(t)\,dt - \int_0^s y(t)\,dt\right| \leqslant \int_0^s |x_n'(t) - y(t)|\,dt$$
$$\leqslant \int_0^1 |x_n'(t) - y(t)|\,dt \leqslant \left(\int_0^1 |x_n'(t) - y(t)|^2\,dt\right)^{\frac{1}{2}}\left(\int_0^1 1^2\,dt\right)^{\frac{1}{2}}.$$

Therefore $\int_0^s x_n'(t)\,dt$ converges to $\int_0^s y(t)\,dt$; the convergence on $[0, 1]$ is uniform with respect to s. In particular, $\int_0^1 y(t)\,dt = 0$, since $\int_0^1 x_n'(t)\,dt = 0$. Next,

$$x_n(0) - x_m(0) = x_n(s) - x_m(s) - \int_0^s [x_n'(t) - x_m'(t)]\,dt$$

$$|x_n(0) - x_m(0)| = \left(\int_0^1 |x_n(0) - x_m(0)|^2\,dt\right)^{\frac{1}{2}}$$
$$\leqslant \left(\int_0^1 |x_n(s) - x_m(s)|^2\,ds\right)^{\frac{1}{2}} + \left\{\int_0^1 \left|\int_0^s [x_n'(t) - x_m'(t)]\,dt\right|^2 ds\right\}^{\frac{1}{2}}.$$

It then follows that $\{x_n(0)\}$ is a Cauchy sequence, with some limit α. Returning to (4.2–1), we see that $x_n(s)$ converges uniformly to the limit $z(s)$, where

$$z(s) = \alpha + \int_0^s y(t)\, dt.$$

Hence $z \in \mathscr{D}(T)$. Also, $z'(s) =^0 y(s)$. The uniform convergence of $x_n(s)$ to $z(s)$ implies that

$$\int_0^1 |x_n(s) - z(s)|^2\, ds \to 0.$$

That is, $x_n \to z$. Since $x_n \to x$, we must have $x = z$. But then $z'(s) =^0 y(s)$ means that $Tx = y$. Hence T is closed.

We now present a group of four very elementary theorems which it is useful to have on record for later use.

Theorem 4.2–B. *Let X be a topological space, and let Y be a Hausdorff space. Suppose f is a continuous function, with range in Y, whose domain is a closed set in X. Then f is closed.*

PROOF. Let $\mathscr{R}$ be the range of f. If we suppose that f is not closed, the complement in $X \times Y$ of the graph is not open, so there must be a point (x_0, y_0) in this complement such that every neighborhood of (x_0, y_0) contains a point of the graph. This implies that every neighborhood of x_0 (in X) contains a point of $\mathscr{D}$ and thus that $x_0 \in \overline{\mathscr{D}}$. But then $x_0 \in \mathscr{D}$, since $\mathscr{D}$ is closed. Then $y_0 \neq f(x_0)$, since otherwise (x_0, y_0) would be in the graph. Since Y is a Hausdorff space, we can find disjoint neighborhoods V_1 and V_2 of $f(x_0)$ and y_0 respectively. From the continuity of f it follows that there is some neighborhood U of x_0 such that $f(x) \in V_1$ if $x \in \mathscr{D} \cap U$. Now $U \times V_2$ is a neighborhood of (x_0, y_0) and must contain a point $(x_1, f(x_1))$ of the graph. Then $x_1 \in \mathscr{D} \cap U$, and so $f(x_1) \in V_1$. But $f(x_1) \in V_2$, and so we reach a contradiction.

Theorem 4.2–C. *If f is closed and f^{-1} exists, f^{-1} is also closed.*

PROOF. The graph of f^{-1} is the set $\{(f(x), x) : x \in \mathscr{D}\}$ in $Y \times X$. The result follows readily from this.

Theorem 4.2–D. *Let X and Y be normed linear spaces, with Y complete. Let A be a linear operator with domain $\mathscr{D} \subset X$ and range $\mathscr{R} \subset Y$. Suppose that A is both closed and continuous. Then $\mathscr{D}$ is closed.*

PROOF. Suppose $x \in \overline{\mathscr{D}}$. Then there exists a sequence $\{x_n\}$ from $\mathscr{D}$ with $x_n \to x$. The sequence $\{Ax_n\}$ is a Cauchy sequence, for $\|Ax_n - Ax_m\| \leqslant C\|x_n - x_m\|$, where C is the norm of A as an operator on the space $\mathscr{D}$.

Hence $\{Ax_n\}$ has some limit $y \in Y$. But then $x \in \mathcal{D}$ and $Ax = y$, since A is closed.

Theorem 4.2–E. *Let X and Y be normed linear spaces, with X complete. Let A be a closed linear operator with domain $\mathcal{D} \subset X$ and range $\mathcal{R} \subset Y$. Suppose that A^{-1} exists and is continuous. Then $\mathcal{R}$ is closed.*

PROOF. A^{-1} is closed (Theorem 4.2–C), and $\mathcal{R}$ is its domain. Hence $\mathcal{R}$ is closed, by Theorem 4.2–D.

There are certain conditions under which it may be concluded that a closed linear operator is continuous. One of the most important theorems of this type is that, if X and Y are complete metric linear spaces (and hence in particular if X and Y are Banach spaces), a linear operator whose domain is all of X and whose range is in Y is continuous if it is closed. This theorem is a consequence of the series of theorems which we shall now consider.

Theorem 4.2–F. *Let X and Y be topological linear spaces. Let T be a linear operator on X into Y, and suppose that the range of T is a set of the second category in Y. Then, to each neighborhood U of 0 in X there corresponds some neighborhood V of 0 in Y such that $V \subset \overline{T(U)}$.*

PROOF. For the convenience of readers who are primarily interested in normed linear spaces, we first give the proof for the case in which X is a normed linear space. Afterwards we shall indicate how to adapt the method of proof for the general case.

If U is a neighborhood of 0 in the normed linear space X, there is an $\epsilon > 0$ such that the open sphere $\|x\| < 2\epsilon$ lies in U. Let W be the open sphere $\|x\| < \epsilon$. Then nW is the open sphere $\|x\| < n\epsilon$, and the union of the sets $T(nW)$ for $n = 1, 2, \cdots$ is the range of T. Consequently, since this range is of the second category in Y, there is some positive integer n such that the interior of $\overline{T(nW)}$ is not empty. But $\overline{T(nW)} = \overline{nT(W)} = n\overline{T(W)}$, and, since $n\overline{T(W)}$ and $\overline{T(W)}$ are homeomorphic sets, the latter also has a nonempty interior. Therefore there exists some point $y_0 \in T(W)$ which is an interior point of $\overline{T(W)}$. It follows that 0 is an interior point of $-y_0 + \overline{T(W)}$. Now let V be a neighborhood of 0 in Y such that $V \subset -y_0 + \overline{T(W)}$. Since $y_0 \in T(W)$, we can write $y_0 = Tx_0$, where $x_0 \in W$. Elements of $-y_0 + T(W)$ have the form $-y_0 + T(w) = T(w - x_0)$, $w \in W$. But $w - x_0 \in U$, because of the way in which W and U are related. Thus $-y_0 + T(W) \subset T(U)$. The closure of $-y_0 + T(W)$ is $-y_0 + \overline{T(W)}$. Therefore $-y_0 + \overline{T(W)} \subset \overline{T(U)}$, and so also $V \subset \overline{T(U)}$. This finishes the proof in this case.

For the case in which X is an arbitrary topological linear space the only difference in the proof concerns the way in which W is chosen. We choose for W a balanced neighborhood of 0 such that $W + W \subset U$ (see Theorem 3.3–E). The fact that W is balanced and absorbing assures that X is the union of the sets nW, $n = 1, 2, \cdots$. Since $-x_0 \in W$ if $x_0 \in W$, it follows in the proof that $w - x_0 \in W + W \subset U$. All the rest of the argument is unchanged.

In the next several theorems we deal with metric linear spaces, as defined in § 3.9. We shall denote the distance from 0 to x by $\|x\|$, rather than by $|x|$, as was done in § 3.9. But the reader must remember that we do not have the property $\|\alpha x\| = |\alpha| \, \|x\|$ at our disposal, since $\|x\|$ need not be a norm. We *do* have $\|-x\| = \|x\|$ and the continuity properties (3.9–6)–(3.9–8). Because we are dealing with metric linear spaces it is necessary, in a few places, to proceed in a manner more roundabout and cumbersome than if we were dealing with normed linear spaces.

Theorem 4.2–G. *Let X and Y be metric linear spaces, and let X be complete. Let T be a closed linear operator with domain $\mathscr{D}$ in X and range $\mathscr{R}$ a set of the second category in Y. For each $\alpha > 0$ let B_α be the closed sphere $\|x\| \leqslant \alpha$ in X and let C_α be the closed sphere $\|y\| \leqslant \alpha$ in Y. We draw the following conclusions:*

1. *To each $\alpha > 0$ corresponds some $\beta > 0$ such that $C_\beta \subset T(\mathscr{D} \cap B_\alpha)$.*
2. *The range $\mathscr{R}$ is all of Y.*
3. *If S is a relatively open subset of $\mathscr{D}$, then $T(S)$ is an open set in Y.*
4. *If T^{-1} exists, it is continuous.*
5. *In particular, if X and Y are normed linear spaces, we can conclude from conclusion 1 that there exists some $m > 0$ such that each $y \in Y$ is of the form $y = Tx$, where $x \in \mathscr{D}$ and $\|x\| \leqslant m\|y\|$.*

PROOF. If $\alpha > 0$ is given, let U be the open sphere $\|x\| < \alpha/2$ in X. Then $U \subset B_{\alpha/2}$. By Theorem 4.2–F there exists an open sphere V in Y, defined by $\|y\| < 2\beta$ for some $\beta > 0$, such that $V \subset \overline{T(\mathscr{D} \cap U)}$ (we take $\mathscr{D}$ for the X in Theorem 4.2–F). Thus certainly $V \subset \overline{T(\mathscr{D} \cap B_{\alpha/2})}$. We shall show that $V \subset T(\mathscr{D} \cap B_\gamma)$ if $\alpha/2 < \gamma$. Choose a sequence $\{\epsilon_n\}$ with

$\epsilon_n > 0$ and $\sum_{1}^{\infty} \epsilon_n < \gamma - \alpha/2$. Let U_n be the open sphere $\|x\| < \epsilon_n$ in X.

We know by Theorem 4.2–F that there exists a positive number δ_n such that $V_n \subset \overline{T(\mathscr{D} \cap U_n)}$ if V_n is the open sphere $\|y\| < \delta_n$ in Y. We can evidently impose on the sequence $\{\delta_n\}$ the further requirement that $\delta_n \to 0$. Now suppose $y \in V$. There exists a point of $T(\mathscr{D} \cap B_{\alpha/2})$, which we can write as $y_0 = Tx_0$, with $x_0 \in \mathscr{D} \cap B_{\alpha/2}$, such that $\|y - y_0\|$

$< \delta_1$. Then $y - y_0 \in V_1$ and hence there exists a point $y_1 = Tx_1$ with $x_1 \in \mathscr{D} \cap U_1$ such that $\|y - y_0 - y_1\| < \delta_2$. Proceeding in this way, by induction we define sequences $\{x_n\}$ and $\{y_n\}$ such that $y_n = Tx_n$, $x_n \in \mathscr{D} \cap U_n$, and $\|y - y_0 - y_1 - \cdots - y_n\| < \delta_{n+1}$, $n = 1, 2, \cdots$. Since $\delta_n \to 0$ we see that $y = \sum_0^\infty y_n$. Now $\|x_n\| < \epsilon_n$ if $n \geqslant 1$, and $\sum_0^\infty \|x_n\| < \sum_1^\infty \epsilon_n < \gamma - \alpha/2$. Since X is complete, the series $\sum_1^\infty x_n$ is convergent, and there is a point x of X defined by $x = \sum_0^\infty x_n$. Moreover, $\|x\| < \|x_0\| + \sum_1^\infty \epsilon_n < \gamma$, since $x_0 \in B_{\alpha/2}$. Thus $x \in B_\gamma$. The fact that T is closed enables us to infer that $x \in \mathscr{D}$ and $Tx = y$ (see Theorem 4.2–A). We have now shown that $V \subset T(\mathscr{D} \cap B_\gamma)$. If we take $\gamma = \alpha$, the fact that $C_\beta \subset V$ now shows that $C_\beta \subset T(\mathscr{D} \cap B_\alpha)$. This proves (1).

To prove (2) observe that, for any $y \in Y$, $(1/n)y \to 0$ as $n \to \infty$, and hence $\|(1/n)y\| < \beta$ if n is large enough. Then $(1/n)y \in C_\beta$, $(1/n)y = Tx$, where $x \in \mathscr{D} \cap B_\alpha$, and $y = T(nx)$, whence $y \in \mathscr{R}$. Thus (2) is proved.

To prove (3) we write $S = \mathscr{D} \cap W$, where W is an open set in X. Suppose $y_0 \in T(S)$, so that $y_0 = Tx_0$, $x_0 \in S$. The set $-x_0 + W$ is a neighborhood of 0, and so contains B_α for some $\alpha > 0$. Then $\mathscr{D} \cap B_\alpha \subset \mathscr{D} \cap (-x_0 + W) = -x_0 + \mathscr{D} \cap W$, and, by (1), $C_\beta \subset T(\mathscr{D} \cap B_\alpha) \subset T(-x_0 + \mathscr{D} \cap W) = -y_0 + T(S)$. It follows that $y_0 + C_\beta \subset T(S)$. Since $y_0 + C_\beta$ contains a neighborhood of y_0, $T(S)$ is open. Thus (3) is proved.

The proof of (4) follows at once from (3) by Theorem 2.12–A, because $(T^{-1})^{-1} = T$. For proof of (5), which is concerned with the particular case in which X and Y are normed linear spaces, observe that, if $y \in Y$ and $y \neq 0$, then $\|\beta y/\|y\|\,\| = \beta$. Hence, by (1), there exists $u \in \mathscr{D}$ with $\|u\| \leqslant \alpha$ and $Tu = \beta y/\|y\|$. Let $x = (\|y\|/\beta)u$. Then $Tx = y$ and $\|x\| \leqslant \alpha\|y\|/\beta$. Thus we can take $m = \alpha/\beta$.

Part 3 of the conclusion of Theorem 4.2–G is sometimes referred to as "*the interior-mapping principle.*"

Several important theorems follow readily from Theorem 4.2–G. We cite two such theorems, which will be used repeatedly in later parts of this book.

Theorem 4.2–H. *Let X and Y be complete metric linear spaces. Let T be a linear operator whose domain is X and whose range is all of Y. Suppose that T is continuous and that T^{-1} exists. Then T^{-1} is continuous.*

PROOF. T is closed, by Theorem 4.2–B. The range of T is of the second category in Y, since Y is complete. The conclusion now follows from Theorem 4.2–G.

Theorem 4.2–I. *Let X and Y be complete metric linear spaces. Let T be a closed linear operator whose domain is all of X and whose range is in Y. Then T is continuous.*

PROOF. Form the product space $X \times Y$. It becomes a complete metric linear space if we define

$$\|(x, y)\| = \|x\| + \|y\|, \qquad (x, y) \in X \times Y.$$

Since T is closed, its graph is a closed linear manifold in $X \times Y$ and can therefore be regarded as a complete metric linear space by itself. We define a linear operator A, with the graph of T as its domain and with X as its range, as follows:

$$A(x, Tx) = x.$$

Since $\|x\| \leqslant \|(x, Tx)\|$, A is continuous at $(0, 0)$ and hence at every point of its domain. Evidently A^{-1} exists and is defined by $A^{-1}x = (x, Tx)$. Then A^{-1} is continuous, by Theorem 4.2–H. It follows that T is continuous.

Theorem 4.2–I has many applications and is used in many arguments. It is commonly referred to as the "*closed-graph theorem.*"

We now consider some concrete applications of the last two theorems.

Example 3. Let Y be the Banach space $C[a, b]$, and let X be a linear space consisting of those functions in $C[a, b]$ which have continuous first and second derivatives on $[a, b]$ and satisfy some definitely specified linear end conditions (e.g., $x(a) = 0$ and $x'(a) = 0$, or $x(a) = x(b) = 0$). Let T be the linear differential operator with domain X and range in Y defined by $Tx = y$, where

$$a_0(s)x''(s) + a_1(s)x'(s) + a_2(s)x(s) = y(s). \tag{4.2–2}$$

The fixed coefficients a_0, a_1, a_2 are assumed to be in Y. We can make X into a Banach space by defining $\|x\|$ to be the greatest of the maximum values on $[a, b]$ of $|x(s)|$, $|x'(s)|$, $|x''(s)|$, respectively. The operator T is continuous, for if

$$M = \|a_0\| + \|a_1\| + \|a_2\|,$$

where $\|a_k\|$ is the norm of a_k as an element of Y, we see that $\|y\| = \|Tx\| \leqslant M\|x\|$.

Now let us suppose that the differential equation (4.2–2) has a unique solution $x \in X$ (i.e., satisfying the specified end conditions) for each choice of $y \in Y$. That is, we assume $\mathscr{R}(T) = Y$ and that T^{-1} exists. It then follows by Theorem 4.2–H that T^{-1} is continuous. The continuous dependence of x on y offers a certain assurance that "perturbation-theory" methods for approximating the solution of the differential equation are satisfactory. That is, a small perturbation of the function y *will result in a small perturbation of the solution* x *and of its first and* second derivatives. The striking thing about this result is that it is obtained without any detailed knowledge of the behavior of the coefficients a_0, a_1, a_2 or of the nature of the end conditions. There is, of course, the strong assumption that the problem always has a unique solution.

The principle involved in the foregoing example can be applied to partial differential equations also.

Example 4. Suppose $1 < p < \infty$, and suppose a sequence of scalars $\alpha_1, \alpha_2, \alpha_3, \cdots$ is given, with the assumption that the series

$$\sum_{1}^{\infty} \alpha_k \xi_k$$

is convergent whenever $x = \{\xi_k\}$ is an element of the space l^p. We shall prove that the sequence $\{\alpha_k\}$ belongs to $l^{p'}$, where $p' = p/(p-1)$.

For each x and i let

$$\eta_i = \sum_{k=1}^{i} \alpha_k \xi_k \qquad i = 1, 2, \cdots. \tag{4.2–3}$$

The sequence $\{\eta_i\}$ is bounded, since it is convergent. Hence equations (4.2–3) can be regarded as defining a linear mapping T of l^p into l^∞, where $y = \{\eta_i\}$ and $y = Tx$. This operator T is closed, as we shall see. For, suppose $x_n \to x$ and $Tx_n \to z$, where x_n and x are in l^p and $z \in l^\infty$. Let $z = \{\zeta_i\}$ and $x_n = \{\xi_k^{(n)}\}$. Then $Tx_n \to z$ means that $\sum_{k=1}^{i} \alpha_k \xi_k^{(n)}$ converges uniformly to ζ_i as $n \to \infty$. Now, by Hölder's inequality,

$$\begin{aligned} \left| \sum_{k=1}^{i} \alpha_k \xi_k^{(n)} - \sum_{k=1}^{i} \alpha_k \xi_k \right| &\leqslant \left(\sum_{k=1}^{i} |\xi_k^{(n)} - \xi_k|^p \right)^{1/p} \left(\sum_{k=1}^{i} |\alpha_k|^{p'} \right)^{1/p'} \\ &\leqslant \|x_n - x\| \left(\sum_{k=1}^{i} |\alpha_k|^{p'} \right)^{1/p'}. \end{aligned}$$

Hence, since $x_n \to x$, the uniqueness of limits shows that

$$\zeta_i = \sum_{k=1}^{i} \alpha_k \xi_k.$$

In other words, $Tx = z$. Thus T is closed. But then it is continuous, by the closed-graph theorem. It then follows that for any x and all i we have

$$\left|\sum_{k=1}^{i} \alpha_k \xi_k\right| \leqslant \|T\| \left(\sum_{k=1}^{\infty} |\xi_k|^p\right)^{1/p}. \tag{4.2–4}$$

Now let us make a special choice of x as follows:

$$\xi_k = \bar{\alpha}_k |\alpha_k|^{p'-2} \qquad \text{if } 1 \leqslant k \leqslant i \text{ and } \alpha_k \neq 0,$$
$$\xi_k = 0 \qquad \text{otherwise.}$$

Then

$$\alpha_k \xi_k = |\xi_k|^p = |\alpha_k|^{p'} \qquad \text{if } 1 \leqslant k \leqslant i.$$

[Observe that $(p' - 1)p = p'$]. When these results are placed in (4.2–4), we have

$$\sum_{k=1}^{i} |\alpha_k|^{p'} \leqslant \|T\| \left(\sum_{k=1}^{i} |\alpha_k|^{p'}\right)^{1/p},$$

whence

$$\left(\sum_{k=1}^{i} |\alpha_k|^{p'}\right)^{1/p'} \leqslant \|T\|.$$

This holds for all i, so it follows that $\{\alpha_k\}$ belongs to $l^{p'}$.

Example 5. Suppose $1 \leqslant p \leqslant \infty$ and $1 \leqslant q \leqslant \infty$. Let $\{\alpha_{ij}\}$ be an infinite matrix $(i, j = 1, 2, \cdots)$ of scalars with the following property: For each $x = \{\xi_j\}$ in l^p the series

$$\eta_i = \sum_{j=1}^{\infty} \alpha_{ij} \xi_j,$$

is convergent for each i, and the sequence $y = \{\eta_i\}$ belongs to l^q. Then the operator A defined by $y = Ax$ is a continuous linear mapping of l^p into l^q.

To prove this we observe first of all that

$$x_i'(x) = \sum_{j=1}^{\infty} \alpha_{ij} \xi_j$$

defines a continuous linear functional x_i' on l^p. If $1 < p < \infty$, this follows from the result established in Example 4, for

$$|x_i'(x)| \leqslant \left(\sum_{j=1}^{\infty} |\alpha_{ij}|^{p'}\right)^{1/p'} \left(\sum_{j=1}^{\infty} |\xi_j|^p\right)^{1/p},$$

so that x_i' is continuous, with

$$\|x_i'\| \leqslant \left(\sum_{j=1}^{\infty} |\alpha_{ij}|^{p'}\right)^{1/p'} < \infty.$$

(Actually, the last $\leqslant$ can be replaced by $=$, but we do not need this fact here.) If $p = 1$ we have

$$\|x_i'\| \leqslant \sup_j |\alpha_{ij}| < \infty;$$

if $p = \infty$ we have

$$\|x_i'\| \leqslant \sum_{j=1}^{\infty} |\alpha_{ij}| < \infty.$$

These latter facts result from problems 4 and 5 at the end of this section. Now, to prove that A is continuous, it suffices to prove that it is closed, by Theorem 4.2–I. Suppose that $x_n \to x$ and $Ax_n \to y$, where x_n, $x \in l^p$ and $y \in l^q$. From $x_n \to x$ we conclude $x_i'(x_n) \to x_i'(x)$ as $n \to \infty$, since x_i' is continuous. Now, if $z = \{\zeta_i\}$ is any member of l^q, ζ_i is a continuous function of z, as we see from the fact that $|\zeta_i| \leqslant \|z\|$. Since $Ax_n = \{x_i'(x_n)\}$ and $y = \{\eta_i\}$, we conclude from $Ax_n \to y$ that $x_i'(x_n) \to \eta_i$. But then $\eta_i = x_i'(x)$. This means that $y = Ax$. Thus A is closed, and the argument is complete.

PROBLEMS

1. Suppose that X is a linear space in which two invariant metrics are defined, in such a way that X is a complete metric linear space with respect to each metric. Suppose that the two topologies are $\mathscr{T}_1$, $\mathscr{T}_2$, and that $\mathscr{T}_1$ is stronger than $\mathscr{T}_2$ (or, equivalently, that if $x_n \to x$ in the sense of $\mathscr{T}_1$, then $x_n \to x$ in the sense of $\mathscr{T}_2$). Then, the two topologies are actually the same. Use Theorem 4.2–H and the identity mapping of X on itself, with the topology $\mathscr{T}_1$ for the domain and $\mathscr{T}_2$ for the range.

2. Let X be a linear space with two norms $\|x\|_1$, $\|x\|_2$, and let X be complete with respect to each norm. Suppose there exists M such that $\|x\|_2 \leqslant M\|x\|_1$ for every x. Then, there is also an m such that $\|x\|_1 \leqslant m\|x\|_2$.

3. Prove Theorem 4.2–I by defining another invariant metric on X, with $\|x\|_1 = \|x\| + \|Tx\|$. Then use problem 1.

4. If $\sum_1^\infty \alpha_k\xi_k$ is convergent whenever $\sum_1^\infty |\xi_k| < \infty$, then $\sup_k |\alpha_k| < \infty$.

5. If $\sum_1^\infty \alpha_k\xi_k$ is convergent whenever $\lim_{k\to\infty} \xi_k = 0$, then $\sum_1^\infty |\alpha_k| < \infty$.

6. For each $x \in L(0, 2\pi)$ let

$$\xi_n = \frac{1}{2\pi}\int_0^{2\pi} x(t)e^{-int}\,dt, \qquad n = 0, \pm 1, \pm 2, \cdots.$$

Suppose X is some closed subspace of $L(0, 2\pi)$ such that $\sum_{-\infty}^{\infty} |\xi_n| < \infty$ for each $x \in X$. Show that there is some constant M such that

$$\sum_{-\infty}^{\infty} |\xi_n| \leqslant M\int_0^{2\pi} |x(t)|\,dt$$

for each $x \in X$.

7. Suppose $x \in L(0, 2\pi)$ and

$$y(t) = \frac{1}{2\pi}\int_0^{2\pi} \frac{x(\theta)}{1 - te^{-i\theta}}\,d\theta,$$

where t is complex and $|t| < 1$. The two following propositions are true (A. E. Taylor, **4**, p. 41 and p. 45):

a. If $1 < p < \infty$ and $x \in L^p(0, 2\pi)$, then $y \in H^p$ (see Example 8, § 3.11 for the definition of H^p).

b. There exists a constant M_p, depending only on p, such that

$$\left(\frac{1}{2\pi}\int_0^{2\pi} |y(re^{i\theta})|^p\,d\theta\right)^{1/p} \leqslant M_p\left(\int_0^{2\pi} |x(\theta)|^p\,d\theta\right)^{1/p}$$

for all $x \in L^p(0, 2\pi)$ and all r, $0 \leqslant r < 1$.

Show that b is a consequence of a, by means of the closed-graph theorem.

8. Prove that the operator A of Example 4, § 1.5, is closed, with domain and range in $L(0, \infty)$. It suffices to deal with the case when $\lambda = 0$, since the addition of a scalar multiple of I to a closed operator leaves a closed operator.

4.3 The Normed Conjugate of a Normed Linear Space

Let X be a normed linear space, and let A be the associated scalar field. This field is itself a Banach space, with the absolute value $|\alpha|$ as the norm of α. A continuous linear operator on X into A, that is, a continuous linear functional on X, is an element of the space $[X, A]$, as defined in § 4.1. We shall denote this space by X'. It is a subspace of X^f. The space X' becomes a Banach space if

$$\|x'\| = \sup_{\|x\|\leqslant 1} |x'(x)|. \tag{4.3-1}$$

When X' is normed in this way, we call it the *normed conjugate* of X.

Some authors call X' the *dual* of X; others call it the *adjoint* of X. Sometimes it is denoted by X^* or by $\bar{X}$.

Our first concern is to establish the fact that X' has sufficiently many elements to meet our needs in some situations which will soon arise. For what follows the reader will need to refer back to § 3.7.

Theorem 4.3–A. *Suppose M is a proper subspace of the normed linear space X. If $m' \in M'$, X' contains an element x' such that $\|x'\| = \|m'\|$ and $x'(x) = m'(x)$ if $x \in M$.*

PROOF. We define $p(x) = \|m'\| \, \|x\|$, $x \in X$. Then p is a norm on X. We observe that $|m'(x)| \leqslant p(x)$ if $x \in M$. We apply Theorem 3.7–B, taking m' and x' in place of the f and F of the theorem referred to. Then $|x'(x)| \leqslant \|m'\| \, \|x\|$ for each x in X. This implies that $x' \in X'$ and $\|x'\| \leqslant \|m'\|$. But $\|m'\| \leqslant \|x'\|$, as a result of the fact that x' is an extension of m'. Thus all is proved.

Theorem 4.3–A is commonly called the Hahn-Banach theorem.

A number of simple but important consequences follow rapidly from Theorem 4.3–A.

Theorem 4.3–B. *If X is a normed linear space, and $x_0 \in X$ with $x_0 \neq 0$, then there exists $x' \in X'$ such that $\|x'\| = 1$ and $x'(x_0) = \|x_0\|$. As a consequence we have*

$$\sup_{\|x'\|=1} |x'(x)| = \|x\| \tag{4.3–2}$$

for each $x \in X$.

PROOF. Let M be the subspace of X generated by x_0. If $x = \alpha x_0$, define $m'(x) = \alpha\|x_0\|$. It is clear that $m' \in M'$ and $\|m'\| = 1$. The existence of the required x' now follows from Theorem 4.3–A. Verification of (4.3–2) is left to the reader.

Theorem 4.3–C. *Suppose x is an element of the normed linear space X such that $x'(x) = 0$ for every $x' \in X'$. Then $x = 0$.*

PROOF. Apply (4.3–2).

Theorem 4.3–D. *Let X be a normed linear space, and let X_0 be a subspace of X. Suppose there exists an x_1 in X at a positive distance h from X_0. (This will be the case if X_0 is closed and a proper subset of X, in which case x_1 may be any element of $X - X_0$.) Then, there exists $x' \in X'$ such that $x'(x_1) = h$, $\|x'\| = 1$, and $x'(x) = 0$ if $x \in X_0$.*

PROOF. First we give a demonstration, based on Theorem 4.3–A. This argument is most convenient for readers interested primarily in normed linear spaces, if they have not studied §§ 3.5, 3.6 in detail. After that we show the relation of Theorem 4.3–D to Theorem 3.6–E and in so doing indicate an alternative proof.

Let M be the subspace generated by X_0 and x_1. Elements of M are representable in the form $m = \alpha x_1 + x$, $x \in X_0$; α and x are uniquely determined by m, because x_1 is not in X_0. Let us define m' by $m'(m) = \alpha h$. We shall show that $m' \in M'$ and $\|m'\| = 1$. If $\alpha \neq 0$ we have $\|m\| = \|\alpha x_1 + x\| = \| -\alpha(-\alpha^{-1}x - x_1)\| \geqslant |\alpha|h$, because $-\alpha^{-1}x \in X_0$ and the distance from x_1 to X_0 is h. Thus $|m'(m)| = |\alpha|h \leqslant \|m\|$ if $\alpha \neq 0$. This inequality is obviously true if $\alpha = 0$. Consequently $m' \in M'$ and $\|m'\| \leqslant 1$. Now, if $\epsilon > 0$ there exists $x \in X_0$ such that $\|x - x_1\| < h + \epsilon$. Let

$$y = \frac{x - x_1}{\|x - x_1\|}.$$

Then $y \in M$, $\|y\| = 1$, and

$$|m'(y)| = \frac{h}{\|x - x_1\|} > \frac{h}{h + \epsilon}.$$

Consequently $\|m'\| > \dfrac{h}{h + \epsilon}$. Since $h \neq 0$ and this is true for every $\epsilon > 0$, we conclude that $\|m'\| \geqslant 1$. Observe that $m'(x_1) = h$ and that $m'(x) = 0$ if $x \in X_0$. The existence of the required $x' \in X'$ now follows from Theorem 4.3–A.

To show the relation of Theorem 4.3–D to Theorem 3.6–E, let $K = \{x : \|x - x_1\| < h\}$. Then K and X_0 do not intersect. Theorem 3.6–E shows that there exists a closed hyperplane which contains X_0 and is such that K lies strictly on one side of it. Since $0 \in X_0$, by Theorem 3.5–E the hyperplane is the set $\{x : x'(x) = 0\}$, where x' is some element of X'. By replacing x' with some constant multiple of itself, we can arrange matters so that $x'(x_1)$ has any specified nonzero value. We choose to make $x'(x_1) = h$. Then $x'(x) > 0$ if $x \in K$. The remainder of the proof consists in showing that $\|x'\| = 1$. Suppose that y is any vector for which $\|y\| < 1$. Then $x_1 \pm hy \in K$, and so $0 < x'(x_1 \pm hy) = h \pm hx'(y)$. From this we obtain $-1 < x'(y) < 1$. It follows that $|x'(x)| \leqslant 1$ if $\|x\| \leqslant 1$. Hence $\|x'\| \leqslant 1$. On the other hand, we can choose a sequence $\{y_n\}$ from X_0 so that $\|y_n - x_1\| \to h$. Let $z_n = \|y_n - x_1\|^{-1}(y_n - x_1)$. Then $\|z_n\| = 1$ and $x'(z_n) = -h\|y_n - x_1\|^{-1}$, so that $|x'(z_n)| \to 1$. Hence $\|x'\| \geqslant 1$, and the proof is complete.

The next two theorems show some uses of the Hahn-Banach theorem and its immediate consequences.

Theorem 4.3–E. *If X' is separable, so is X.*

PROOF. Let $\{x_n'\}$ be a countable set which is everywhere dense on the set $\{x' : \|x'\| = 1\}$ in X'. Choose $x_n \in X$ so that $\|x_n\| = 1$ and $|x_n'(x_n)| \geqslant {}^3/_4$. Let M be the closed linear manifold in X generated by the sequence $\{x_n\}$. Suppose $M \neq X$ and $x_0 \in X - M$. Then Theorem 4.3–D tells us that there exists $x' \in X'$ such that $\|x'\| = 1$, $x'(x_0) \neq 0$, and $x'(x) = 0$ if $x \in M$. Then $x'(x_n) = 0$ if $n = 1, 2, \cdots$, and

$$\tfrac{3}{4} \leqslant |x_n'(x_n)| \leqslant |x_n'(x_n) - x'(x_n)| + |x'(x_n)|,$$

whence

$$\tfrac{3}{4} \leqslant \|x_n' - x'\| \, \|x_n\| = \|x_n' - x'\|.$$

This contradicts the fact that $\|x_n' - x'\|$ can be made as small as we please by suitable choice of n. Hence $M = X$. It then follows that linear combinations formed from $\{x_n\}$ with rational scalar coefficients constitute a countable set everywhere dense in X, so that X is separable.

We shall see that X can be separable and X' not, so that the converse of Theorem 4.3–E is false. For example, l^1 is separable, but it turns out (see Theorem 4.32–A) that $(l^1)'$ is congruent to l^∞, which is not separable.

The next theorem involves the concept of a quotient space as a normed linear space (see § 3.14).

Theorem 4.3–F. *Let X be a normed linear space and M a subspace of it. Let $M^0 = \{x' \in X' : x'(x) = 0$ if $x \in M\}$. Then M^0 is a closed subspace of X', and the conjugate space M' is congruent to X'/M^0.*

PROOF. It is evident that M^0 is a closed subspace. In what follows we use m and m' for typical elements of M and M', respectively. A typical element of X'/M^0 is $[x']$, where $x' \in X'$. We have $y' \in [x']$ if and only if $y' - x' \in M^0$, i.e., $y'(m) = x'(m)$ for every m. With $[x']$ given we define m' by $m'(m) = y'(m)$, where $y' \in [x']$. Clearly m' depends only on $[x']$ and not on the particular y' chosen. Also, $m' \in M'$. Now $|m'(m)| \leqslant \|y'\| \, \|m\|$, and so, by the way in which $\|[x']\|$ is defined, we conclude that $\|m'\| \leqslant \|[x']\|$. On the other hand, Theorem 4.3–A shows that there exists some $y' \in [x']$ with $\|y'\| = \|m'\|$. Therefore

$$\|[x']\| = \inf_{y' \in [x']} \|y'\| \leqslant \|m'\|.$$

Thus $\|m'\| = \|[x']\|$. If we write $m' = T([x'])$, we see that T defines a linear and isometric mapping of X'/M^0 into M'. The range of T is in fact all of M', for if $m' \in M'$, and $x' \in X'$ is such that $x'(m) = m'(m)$ when $m \in M$, we have $m' = T([x'])$. Such an x' exists, by Theorem 4.3–A. This completes the proof.

The Hahn-Banach theorem can be used to prove an existence theorem for Green's function for the Dirichlet problem, under certain conditions.

We present a brief sketch of the line of argument of this proof. For simplicity we consider the two-dimensional case.

Let D be a bounded, connected open set in the plane, with boundary C consisting of a finite number of smooth curves. Let B be the real Banach space of continuous functions f defined on C, with $\|f\|$ = the maximum value of $|f|$ on C. Let B_0 be the linear manifold in B consisting of those f for which the Dirichlet problem for the region D is solvable (see the discussion of Example 1, § 1.5). With each point $Q \in D$ we associate a continuous linear functional l defined on B_0, the value of l at $f \in B_0$ being $u(Q)$, where u is the solution of the Dirichlet problem for the boundary-value function f. That is, u is harmonic in D, continuous in $\bar{D}$, and $u = f$ on C. This functional l is linear. Since $u \equiv 1$ if $f \equiv 1$, and since $|u(Q)| \leqslant$ maximum of $|f|$ on C (by the maximum-value theorem for harmonic functions), it is evident that $l \in B_0'$ and $\|l\| = 1$. By Theorem 4.3–A there exists an element L of B' which is an extension of l and for which $\|L\| = 1$. We shall denote L by L_Q to exhibit its dependence on Q.

Now let P be any point not on C in the plane. If t represents a point on C, let $g_P(t) = \log \overline{tP}$. Note that $g_P \in B_0$ if P is in the complement of $\bar{D}$. The value at Q of the corresponding solution of the Dirichlet problem is $u(Q) = l(g_P) = \log \overline{QP}$. Now, for any fixed P not on C, g_P is an element of B, and so we can apply L_Q to g_P; we define

$$k(P, Q) = L_Q(g_P).$$

If P is on C (and Q is in D), we define

$$k(P, Q) = \log \overline{QP}.$$

We then define

$$G(P, Q) = -\log \overline{QP} + k(P, Q)$$

for $Q \in D$ and any P. We assert that G is Green's function (to be considered as a function of P with singularity, or "pole," at Q). To show this it is merely necessary to prove that, as a function of P, $k(P, Q)$ is continuous on the set $\bar{D}$ and harmonic in D. Let Δ_P denote the Laplacian operator. Then, for P in D,

$$\Delta_P k(P, Q) = L_Q(\Delta_P g_P) = L_Q(0) = 0.$$

The commuting of Δ_P and L_Q is justified by the fact that L_Q is a bounded operator, just as in differentiating under an integral sign. Hence $k(P, Q)$ is harmonic in D. All that remains is to prove continuity of $k(P, Q)$ at points P on C.

We know that $L_Q(g_P) = \log \overline{QP}$ if $Q \in D$ and P is *not* in $\bar{D}$. Hence, if $P_0 \in C$ and $R \to P_0$ from outside of $\bar{D}$, it follows that $k(R, Q) \to \log \overline{QP_0} = k(P_0, Q)$. We want to prove that $k(P, Q) \to k(P_0, Q)$ as $P \to P_0$ from inside C. It will suffice to show that, with each point P sufficiently near P_0 and in D, we can associate a point R not in $\bar{D}$ such that $R \to P_0$ and $k(P, Q) - k(R, Q) \to 0$ as $P \to P_0$. We demonstrate this as follows. With P given, draw a straight line from P to a nearest point N of C; this line is normal to C at N. Continue the line beyond N to a point R such that $\overline{PN} = \overline{NR}$. If P is sufficiently near P_0, R will not be in $\bar{D}$. We have

$$k(P, Q) - k(R, Q) = L_Q(g_P - g_R),$$

$$g_P(t) - g_R(t) = \log \frac{\overline{tP}}{\overline{tR}}.$$

It can be proved that, as $P \to P_0$, $R \to P_0$ and

$$\frac{\overline{tP}}{\overline{tR}} \to 1$$

uniformly with respect to t on C. Thus $\|g_P - g_R\| \to 0$. The desired result now follows, owing to the continuity of the linear functional L_Q.

For more details on this subject, see the papers by P. D. Lax and C. Miranda listed in the bibliography.

If X is an incomplete normed linear space, and $\hat{X}$ is the completion of X (see § 3.13), it turns out that X' and $\hat{X}'$ are congruent, and it is convenient to regard X' and $\hat{X}'$ as being the same space. The fact that they are congruent follows from Theorem 3.13–A, with Y the space of scalars.

PROBLEMS

1. If X is a normed linear space of finite dimension n, the dimension of X' is also n. Prove this by showing that $X' = X^f$. See Theorem 1.61–A and the proof of Theorem 3.12–A.

2. The proposition in (1) can be generalized. If X is an n-dimensional topological linear T_1-space, all the elements of X^f are continuous, and X^f is n-dimensional. See problems 1 and 2 of § 3.3.

3. If X is a complex normed linear space and X_r is the associated real linear space, we write $x'(x) = x_1'(x) - ix_1'(ix)$, where $x_1' \in (X_r)'$ and $x' \in X'$ (see § 3.7). Show that $\|x_1'\| = \|x'\|$.

4. Suppose M is a subspace of the normed linear space X. Let $G = M \cap \{x : \|x\| \leqslant 1\}$. Suppose $x_0' \notin M^0$. Show that the distance from x_0' to M^0 is $\sup_{x \in G} |x_0'(x)|$ and that there is in M^0 an element x' such that $\|x' - x_0'\| = d(x_0', M^0)$.

5. If X is a Banach space of infinite dimension, its dimension (see problem 2, § 1.72) is at least as great as the cardinality C of the set of all real numbers. The proof is made (*a*) by observing that the dimension of l^∞ is at least as great as C, and (*b*) by constructing a one-to-one linear mapping of l^∞ into X. For (*a*) consider the elements $(t, t^2, t^3, \cdots)$ of l^∞, where $0 < t < 1$. For (*b*) construct a sequence $\{x_n\}$ of elements of X and a sequence $\{M_n\}$ of closed subspaces of X such that $\sum_n \|x_n\| < \infty$, $x_n \notin M_n$, $x_k \in M_n$ if $k > n$. Then, if $y = \{\xi_k\} \in l^\infty$, define $Ty = x = \sum_1^\infty \xi_k x_k$; T is the desired mapping of l^∞ into X. To obtain $\{M_n\}$ and $\{x_n\}$ let $M_0 = X$ and for $n \geqslant 1$ choose $m_n' \in M_{n-1}'$ with $m_n' \neq 0$; then define $M_n = \{x \in M_{n-1}: m_n'(x) = 0\}$ and choose $x_n \in M_{n-1} - M_n$ in a suitable way.

6. If X is an infinite-dimensional separable Banach space, its dimension is C. This can be shown by using the results of the foregoing problem and of problem 2, § 1.72, for the cardinality of X cannot exceed C.

7. If X is a Banach space of infinite dimension, the weak topology $\mathscr{T}(X, X')$ for X is not metrizable. Use problem 5 and § 3.9, problem 2.

8. Let X be a normed linear space and M a subspace of X'. If M is not closed, then $\mathscr{T}(X, \bar{M})$ is strictly stronger than $\mathscr{T}(X, M)$ (see § 3.81). However, if S is a bounded subset of X, the relative topologies induced on S by $\mathscr{T}(X, M)$ and $\mathscr{T}(X, \bar{M})$ are the same. To prove this it suffices to show that, if $\epsilon > 0$ and $x_1', \cdots, x_n'$ are in $\bar{M}$, the set $V \cap S$, where $V = \{x: |x_i'(x)| < \epsilon, i = 1, \cdots, n\}$, is open in the topology induced on S by $\mathscr{T}(X, M)$.

4.31 The Second Normed Conjugate Space

If X is a normed linear space, we denote the normed conjugate of X' by X''. We shall use x'' as a typical notation for elements of X'', and we shall often write $\langle x', x''\rangle$ in place of $x''(x')$.

Note that, although X' is a subspace of X^f, X'' is not in general a subspace of X^{ff} (though it *is* a subspace of $(X')^f$). There is, however, a canonical mapping of X into X'' which is similar to the canonical mapping of X into X^{ff}, as defined in § 1.6. We shall use the symbol J for this canonical mapping of X into X'', in spite of the fact that it is not the same as the J of § 1.6. Since we do not consider both mappings at the same time, no confusion need arise.

The definition of the canonical mapping of X into X'' is expressed by the formula

$$\langle x', Jx\rangle = \langle x, x'\rangle, \qquad x \in X, \qquad x' \in X'. \tag{4.31–1}$$

That is, $Jx = x''$, where $x''(x') = x'(x)$. It is evident that $x'' \in (X')^f$, and since $|x''(x')| \leqslant \|x'\|\,\|x\|$, it is clear that $x'' \in X''$ and $\|x''\| \leqslant \|x\|$. Actually, we can prove that $\|x''\| = \|x\|$. This is obviously so if $x = 0$, and so we assume $x \neq 0$. Then, there exists $x' \in X'$ such that $\|x'\| = 1$ and $x'(x) = \|x\|$ (Theorem 4.3–B). By the formula for $\|x''\|$ corresponding to (4.3–1) we see then that $\|x''\| \geqslant \|x\|$. Hence $\|x''\| = \|x\|$. We state this result for reference in the form

$$\|Jx\| = \|x\|. \tag{4.31–2}$$

Evidently J sets up a congruence between X and a subspace of X''.

Definition. If the range of J is all of X'', we say that X is *norm reflexive.* If the exact context is clear, we may use the simpler term *reflexive* instead of *norm reflexive.* But it should be recalled that we have already (in § 1.6) defined a concept of *algebraic reflexivity* for a linear space. In contrast to the situation with algebraic reflexivity, we shall see that there exist infinite-dimensional spaces which are norm reflexive.

Since X'' is complete, whether X is or not, it is clear from (4.31–2) that, if X is incomplete, it cannot be reflexive.

Theorem 4.31–A. *If X is norm reflexive, so is X'.*

PROOF. The argument is very straightforward. Denote by J_0 the canonical mapping of X into X'', and by J_1 that of X' into X'''. We assume $\mathscr{R}(J_0) = X''$, and we wish to prove $\mathscr{R}(J_1) = X'''$. Suppose $x''' \in X'''$, and define x' by $\langle x, x' \rangle = \langle J_0 x, x''' \rangle$, x varying over X. The proof is then completed by proving that $x' \in X'$ and $J_1 x' = x'''$. We leave the last details to the reader.

It will be shown in § 4.6 (problem 4) that, if X is complete and X' is norm reflexive, so is X.

Theorem 4.31–B. *If X is norm reflexive, the same is true of each closed subspace of X.*

PROOF. Let M be a closed subspace of X. We use Theorem 4.3–F and the notation employed in its proof. In particular, we use the operator T which maps X'/M^0 isometrically and linearly onto M'. Suppose $m'' \in M''$. Define $x'' \in X''$ by setting $x''(x') = m''\{T([x'])\}$ for each x' in X'. Since X is reflexive, there exists $x \in X$ such that $x''(x') = x'(x)$ for each x'. We assert that $x \in M$. For, if not, the fact that M is closed enables us to infer that there exists $x' \in X'$ such that $x' \in M^0$ and $x'(x) \neq 0$ (by Theorem 4.3–D). Then $[x']$ is the zero element of X'/M^0, and so $T([x']) = 0$, whence $x'(x) = x''(x') = m''\{T([x'])\} = 0$, contrary to fact. Let us now denote x by m, since $x \in M$. If $m' = T[x']$ is any element of M', we have $m'(m) = x'(m)$, by the way in which T is defined. Thus $m''(m') = x''(x') = x'(m) = m'(m)$. This proves that M is reflexive.

PROBLEMS

1. A finite-dimensional normed linear space is norm reflexive.

2. The range of J is closed in X'' if and only if X is complete.

3. If X is complete but not norm reflexive, $J(X)$ is of the first category in X''.

4.32 Some Representations of Linear Functionals

In this section we shall obtain representation theorems for linear functionals on certain particular normed linear spaces. These representation theorems are of the following sort: If X is a certain given space, we find another normed linear space Y which is congruent to X' in such a way that, if $x' \in X'$ and if the corresponding element of Y is y, for each $x \in X$ the value of $x'(x)$ is expressed in a well-determined way by an analytical process involving x and y. This process usually involves an infinite series or some kind of an integral.

Several important results are given as problems.

THE NORMED CONJUGATE OF l^p, $1 \leqslant p < \infty$

The spaces l^p were discussed in § 3.11. With each p we associate p', defined as follows:

$$\begin{aligned} p' &= \frac{p}{p-1} && \text{if } 1 < p < \infty, \\ p' &= \infty && \text{if } p = 1, \\ p' &= 1 && \text{if } p = \infty. \end{aligned}$$

Note that $(p')' = p$ and that

$$\frac{1}{p} + \frac{1}{p'} = 1 \quad \text{if} \quad 1 < p < \infty. \tag{4.32–1}$$

Note also that $p' = 2$ if and only if $p = 2$, and that $1 < p < 2$ implies $2 < p' < \infty$.

For general use, here and elsewhere, we define

$$\operatorname{sgn} c = \begin{cases} 0 & \text{if } c = 0 \\ \dfrac{c}{|c|} & \text{if } c \neq 0. \end{cases}$$

Here c may be any real or complex number. Sometimes it is convenient in printing to denote $\operatorname{sgn} \bar{c}$ by $\overline{\operatorname{sgn}}\, c$, especially if c is replaced by a lengthy or bulky expression. Observe that $c \operatorname{sgn} \bar{c} = |c|$, and that $|\operatorname{sgn} c| = 0$ or 1 according to whether c is 0 or $\neq 0$.

If $x = \{\xi_n\} \in l^p$, we denote by u_k that x for which $\xi_k = 1$ and $\xi_n = 0$ if $n \neq k$. If $1 \leqslant p < \infty$, we see that for any $x \in l^p$

$$\left\| x - \sum_1^n \xi_k u_k \right\| = \left(\sum_{n+1}^{\infty} |\xi_k|^p \right)^{1/p} \rightarrow 0$$

as $n \rightarrow \infty$, so that

$$x = \sum_1^{\infty} \xi_k u_k. \tag{4.32–2}$$

It follows that

$$x'(x) = \sum_1^{\infty} \xi_k x'(u_k) \tag{4.32–3}$$

if $x' \in X'$. The condition $p < \infty$ is essential for (4.32–2).

The problem of representing x' is now seen to be the problem of describing what sets of values are admissible for $x'(u_k)$, $k = 1, 2, \cdots$, and of giving a formula for $\|x'\|$ in terms of these values. The solution of the problem is given in the following theorem.

Theorem 4.32–A. *Suppose $1 \leqslant p < \infty$. Every continuous linear functional on l^p is representable in one and only one way in the form*

$$x'(x) = \sum_1^{\infty} \alpha_k \xi_k, \tag{4.32–4}$$

where the sequence $a = \{\alpha_n\}$ is an element of $l^{p'}$. Every element a of $l^{p'}$ can be used in this way to define an element of $(l^p)'$, and the correspondence between x' and a is a congruence of $(l^p)'$ and $l^{p'}$. In particular,

$$\|x'\| = \begin{cases} \left(\sum_1^{\infty} |\alpha_k|^{p'} \right)^{1/p'} & \textit{if } 1 < p < \infty \\ \sup_k |\alpha_k| & \textit{if } p = 1. \end{cases} \tag{4.32–5}$$

PROOF. Suppose $x' \in (l^p)'$ given. Define $\alpha_k = x'(u_k)$. Then $x'(x)$ is given by (4.32–3), or (4.32–4). Suppose first that $1 < p < \infty$. The case $p = 1$ is considered later. For a given positive integer n, choose x as follows:

$$\xi_k = \begin{cases} |\alpha_k|^{p'-1} \operatorname{sgn} \bar{\alpha}_k & \text{if } 1 \leqslant k \leqslant n \\ 0 & \text{if } n < k. \end{cases} \tag{4.32–6}$$

Then, if $1 \leqslant k \leqslant n$,

$$\alpha_k \xi_k = |\alpha_k|^{p'} = |\xi_k|^p,$$

and so

$$\|x\| = \left(\sum_1^n |\alpha_k|^{p'} \right)^{1/p}, \qquad x'(x) = \sum_1^n |\alpha_k|^{p'}.$$

But $|x'(x)| \leqslant \|x'\| \, \|x\|$, and so

$$\sum_1^n |\alpha_k|^{p'} \leqslant \|x'\| \left(\sum_1^n |\alpha_k|^{p'} \right)^{1/p}.$$

It follows that

$$\left(\sum_1^n |\alpha_k|^{p'} \right)^{1/p'} \leqslant \|x'\|.$$

Hence $a = \{\alpha_n\}$ is in $l^{p'}$, and $\|a\| \leqslant \|x'\|$.

Suppose, on the other hand, that $a = \{\alpha_n\}$ is a given element of $l^{p'}$. We can define $x' \in (l^p)'$ by (4.32–4). By Hölder's inequality we have

$$|x'(x)| \leqslant \left(\sum_1^\infty |\alpha_k|^{p'} \right)^{1/p'} \left(\sum_1^\infty |\xi_k|^p \right)^{1/p} = \|a\| \, \|x\|,$$

so that $\|x'\| \leqslant \|a\|$. Since $x'(u_k) = \alpha_k$ in this case it follows from the preceding work that $\|a\| \leqslant \|x'\|$. Hence $\|a\| = \|x'\|$, and all is clear when $1 < p < \infty$.

If $p = 1$ we replace (4.32–6) by

$$\xi_k = \begin{cases} \operatorname{sgn} \bar{\alpha}_n & \text{if } k = n \\ 0 & \text{if } k \neq n. \end{cases}$$

Then $\|x\| \leqslant 1$, and (4.32–4) becomes $x'(x) = |\alpha_n|$, so that $|\alpha_n| \leqslant \|x'\| \, \|x\| \leqslant \|x'\|$, whence $a = \{\alpha_n\} \in l^\infty$, with $\|a\| \leqslant \|x'\|$. The rest of the argument for $p = 1$ is left to the reader.

Theorem 4.32–A is not true for the case $p = \infty$. If it were, $(l^\infty)'$ would be congruent to the separable space l^1. Since l^∞ is not separable, this would contradict Theorem 4.3–E.

THE NORMED CONJUGATE OF $C[a, b]$

For the definition of $C[a, b]$ see Example 3, § 1.3 and Example 4, § 3.11. We shall also have to consider $B[a, b]$ (Example 3, § 3.11), of which $C[a, b]$ is a subspace.

Our problem is to discover an analytic representation for all continuous linear functionals on $C[a, b]$. We shall see that to each such functional x' corresponds a function of bounded variation v such that the values of the functional are given by the Riemann-Stieltjes integral

(4.32–7) $$x'(x) = \int_a^b x(t) \, dv(t).$$

Theorem 4.32–B. *Suppose* $x' \in X'$, *where* $X = C[a, b]$. *Then, there exists a function of bounded variation* v, *defined on* $[a, b]$ *and with values in the scalar field associated with* X, *such that* (4.32–7) *is valid for each* x *and also such that the total variation of* v *on* $[a, b]$ *is* $\|x'\|$.

PROOF. Let f be a continuous linear functional on $B[a, b]$ which is an extension of x' and such that $\|f\| = \|x'\|$. Such an f exists, by Theorem 4.3–A. If s is any point of $[a, b]$, let x_s be defined by $x_a(t) \equiv 0$ (for $s = a$) and by

$$x_s(t) = \begin{cases} 1 & \text{if } a \leqslant t \leqslant s \\ 0 & \text{if } s < t \leqslant b \end{cases}$$

in case $a < s \leqslant b$. Then $x_s \in B[a, b]$. We define v by

$$v(s) = f(x_s). \tag{4.32–8}$$

We now proceed to show that v is of bounded variation and that $V(v) \leqslant \|x'\|$, where $V(v)$ is the total variation of v. Consider any partition of $[a, b]$:

$$a = t_0 < t_1 < \cdots < t_n = b.$$

Using the $\overline{\text{sgn}}$ notation explained earlier, we let

$$\epsilon_i = \overline{\text{sgn}}\,[v(t_i) - v(t_{i-1})], \qquad i = 1, \cdots, n,$$

and we construct the function

$$y(t) = \begin{cases} \epsilon_1 & \text{if } t_0 \leqslant t \leqslant t_1 \\ \epsilon_i & \text{if } t_{i-1} < t \leqslant t_i, \qquad i = 2, \cdots, n. \end{cases}$$

Then $y \in B[a, b]$ and $\|y\| \leqslant 1$. We can write

$$y(t) = \sum_{i=1}^{n} \epsilon_i[y_i(t) - y_{i-1}(t)],$$

where $y_k = x_{t_k}$. Thus

$$f(y) = \sum_{i=1}^{n} \epsilon_i[f(y_i) - f(y_{i-1})] = \sum_{i=1}^{n} \epsilon_i[v(t_i) - v(t_{i-1})]$$

$$= \sum_{i=1}^{n} |v(t_i) - v(t_{i-1})|.$$

But $|f(y)| \leqslant \|f\|\,\|y\| \leqslant \|f\| = \|x'\|$. Therefore

$$\sum_{i=1}^{n} |v(t_i) - v(t_{i-1})| \leqslant \|x'\|.$$

This shows that v is of bounded variation, with $V(v) \leqslant \|x'\|$.

Now we proceed to prove the validity of (4.32–7). With the same notation as before for an arbitrary partition, and with any given x in $C[a, b]$, we define $z \in B[a, b]$ as follows:

$$z(t) = \sum_{i=1}^{n} x(t_{i-1})[y_i(t) - y_{i-1}(t)].$$

Then

$$f(z) = \sum_{i=1}^{n} x(t_{i-1})[v(t_i) - v(t_{i-1})]. \tag{4.32–9}$$

Also,

$$|z(t) - x(t)| = \begin{cases} |x(t_0) - x(t)| & \text{if } a \leqslant t \leqslant t_1 \\ |x(t_{i-1}) - x(t)| & \text{if } t_{i-1} < t \leqslant t_i,\, i = 2, \cdots, n. \end{cases}$$

Let Δ denote the partition, and let

$$|\Delta| = \max \{|t_1 - t_0|, \cdots, |t_n - t_{n-1}|\}.$$

Then we see (by the uniform continuity of x) that $\|z - x\| \to 0$ when $|\Delta| \to 0$. Since f is continuous, $f(z) \to f(x)$. But also, it is clear from (4.32–9) and the definition of a Stieltjes integral that

$$f(z) \to \int_a^b x(t)\, dv(t).$$

We therefore conclude that

$$f(x) = \int_a^b x(t)\, dv(t).$$

This is the same as (4.32–7), since f is an extension of x'. It is a standard property of Stieltjes integrals that

$$\left| \int_a^b x(t)\, dv(t) \right| \leqslant \|x\| V(v). \tag{4.32–10}$$

Consequently $\|x'\| \leqslant V(v)$. We already know that $V(v) \leqslant \|x'\|$, and so $V(v) = \|x'\|$.

Theorem 4.32–B makes no assertion about the uniqueness of v, nor does it assert that X' is congruent to some space of which v is a member. We observe in the first place that there is no unique function of bounded variation which makes (4.32–7) true for all $x \in C[a, b]$. For, we can add an arbitrary constant to v without affecting (4.32–7). It is also easy to see that the values of v at its points of discontinuity in the interior of the interval $[a, b]$ can be altered without affecting the value of the integral in (4.32–7). For, if w is a function of bounded variation which is equal to v at a, b and at all points t in the interior of $[a, b]$ at which v is continuous, the approximating sums which define

$$\int_a^b x(t)\,dv(t) \qquad \text{and} \qquad \int_a^b x(t)\,dw(t)$$

can be formed exclusively by use of values of v and w at points of the aforementioned kind, so that the integrals must be equal in value.

In order to obtain a congruence between the space X' and a suitable space of functions of bounded variation, we proceed in the following manner. As in Example 7, § 3.11, we denote by $BV[a, b]$ the Banach space of functions of bounded variation defined on $[a, b]$, with

$$\|v\| = |v(a)| + V(v).$$

We can define an equivalence relation in $BV[a, b]$ by writing $v_1 \sim v_2$ if

$$\int_a^b x(t)\,dv_1(t) = \int_a^b x(t)\,dv_2(t)$$

for each $x \in C[a, b]$. We assert that $v \sim 0$ if and only if $v(a) = v(b)$ and $v(c - 0) = v(c + 0) = v(a)$ if $a < c < b$. To see this we observe first of all that $v \sim 0$ implies

$$0 = \int_a^b dv(t) = v(b) - v(a).$$

Next we observe the following: If $v \in BV[a, b]$ and $a \leqslant c < b$, then

$$\frac{1}{h}\int_c^{c+h} v(t)\,dt \to v(c + 0) \tag{4.32–11}$$

as $h \to 0$. Likewise, if $a < c \leqslant b$,

$$\frac{1}{h}\int_{c-h}^{c} v(t)\,dt \to v(c - 0) \tag{4.32–12}$$

as $h \to 0$. We leave the simple proofs of these facts to the reader. Now suppose that $a \leqslant c < b$, $0 < h < b - c$, and define

$$x(t) = \begin{cases} 1, & a \leqslant t \leqslant c \\ 1 - \dfrac{t - c}{h}, & c \leqslant t \leqslant c + h \\ 0, & c + h \leqslant t \leqslant b. \end{cases}$$

Then $v \sim 0$ implies

$$0 = \int_a^b x(t)\,dv(t) = v(c) - v(a) + \int_c^{c+h} x(t)\,dv(t).$$

On integration by parts we obtain

$$\int_c^{c+h} x(t)\,dv(t) = -\,v(a) + \frac{1}{h}\int_c^{c+h} v(t)\,dt.$$

Using (4.32–11), we obtain the result $v(c + 0) = v(a)$. A similar argument shows that $v(c - 0) = v(b)$ if $a < c \leqslant b$. [We remark that $v(a + 0) = v(a)$ is a consequence of $v(c + 0) = v(a)$ for $a < c < b$, because there exists a sequence c_n such that $a < c_n < b$, $c_n \to a$, and $v(c_n + 0) = v(c_n)$; likewise for $v(b - 0) = v(b)$.]

Conversely, it is evident that, if $v \in BV[a, b]$ and $v(a) = v(b) = v(c + 0) = v(c - 0)$ when $a < c < b$, then $v \sim 0$. For then $v(t) = v(a)$ at $t = a$, $t = b$, and all interior points of $[a, b]$ at which v is continuous, so that, if $x \in C[a, b]$,

$$\int_a^b x(t)\,dv(t) = \int_a^b x(t)\,dw(t) = 0,$$

where $w(t) \equiv v(a)$.

Now let $X = C[a, b]$. We wish to associate with each $x' \in X'$ a unique $v \in BV[a, b]$ in such a way that the association makes X' congruent to a subspace of $BV[a, b]$, and such that x' and v are related by the formula (4.32–7). We do this by introducing the concept of a normalized function of bounded variation. We shall say that v is *normalized* if $v(a) = 0$ and if $v(t + 0) = v(t)$ when $a < t < b$. Other definitions of normalization can be used; we might alternatively require $v(t - 0) = v(t)$, or

$$v(t) = \tfrac{1}{2}[v(t + 0) + v(t - 0)].$$

Our choice is governed by consideration of some future applications.

The normalized functions form a linear manifold in $BV[a, b]$. Each equivalence class contains at most one normalized function. These facts are readily evident. From any $v \in BV[a, b]$ we can define another member v^* of $BV[a, b]$ as follows:

$$\begin{aligned} &v^*(a) = 0, && v^*(b) = v(b) - v(a) \\ &v^*(t) = v(t + 0) - v(a), && a < t < b. \end{aligned}$$

Then one can easily verify that v^* is normalized and that $v^* \sim v$. Thus each equivalence class contains *exactly* one normalized function. Finally, one may show that $V(v^*) \leqslant V(v)$. For, suppose $a = t_0 < t_1 < \cdots < t_n = b$. If $\epsilon > 0$ we can choose points $s_1, \cdots, s_{n-1}$ at which v is continuous, with s_k so close to t_k (on the right) that

$$|v(t_k + 0) - v(s_k)| < \epsilon/2n.$$

It is then readily seen that, if $s_0 = a$ and $s_n = b$, we have

$$\sum_{k=1}^{n} |v^*(t_k) - v^*(t_{k-1})| \leqslant \sum_{k=1}^{n} |v(s_k) - v(s_{k-1})| + \epsilon \leqslant V(v) + \epsilon,$$

and so $V(v^*) \leqslant V(v) + \epsilon$. But then $V(v^*) \leqslant V(v)$.

We can now state:

Theorem 4.32–C. *Let $X = C[a, b]$. Then X' is congruent to the subspace of $BV[a, b]$ consisting of all normalized functions of bounded variation. If v is such a normalized function, the corresponding x' is given by*

$$x'(x) = \int_a^b x(t)\, dv(t). \tag{4.32–13}$$

PROOF. Formula (4.32–13) defines a linear mapping $x' = Tv$, where v is normalized and $x' \in X'$. We evidently have $\|x'\| \leqslant V(v)$ [see (4.32–10)]. For a normalized v, $V(v)$ is the norm of v, because $v(a) = 0$. Now consider any $x' \in X'$. Theorem 4.32–B tells us that there is some $u \in BV[a, b]$ such that

$$x'(x) = \int_a^b x(t)\, du(t)$$

and $V(u) = \|x'\|$. The integral is not changed if we replace u by u^*. Then $x' = Tu^*$ and $\|x'\| \leqslant V(u^*)$. Also, $V(u^*) \leqslant V(u) = \|x'\|$. Therefore $\|x'\| = V(u^*)$. Since there is just one normalized function in each equivalance class, we see that the theorem is proved.

The following theorem is of interest, especially in relation to § 7.5.

Theorem 4.32–D. *Suppose the functional x' is defined by formula* (4.32–7), *where v is of bounded variation. Suppose also that $x'(x) \geqslant 0$ if $x \in C[a, b]$ and $x(t) \geqslant 0$ for all values of t on $[a, b]$. Then $v(a) \leqslant v(s + 0) \leqslant v(b)$ if $a \leqslant s < b$ and $v(s_1 + 0) \leqslant v(s_2 + 0)$ if $a \leqslant s_1 < s_2 < b$. In particular, v is nondecreasing if it is normalized.*

PROOF. We shall not go into full detail with the argument. The fundamental idea of the proof rests on the following consideration: If $a \leqslant s < b$ and $0 < h < b - s$, let x be the function defined by $x(t) = 1$ if $a \leqslant t \leqslant s$, $x(t) = (s + h - t)/h$ if $s \leqslant t \leqslant s + h$, and $x(t) = 0$ if $s + h \leqslant t \leqslant b$. Then $x'(x) \geqslant 0$. A rather easy calculation shows that, as $h \to 0$,

$$\int_a^b x(t)\, dv(t) \to v(s + 0) - v(a).$$

Thus $v(a) \leqslant v(s + 0)$. The rest of the proof is left to the reader.

PROBLEMS

1. Let (c) denote the space of sequences $x = \{\xi_n\}$ such that $\lim_{n\to\infty} \xi_n = \xi_0$ exists. Let $\|x\| = \sup_n |\xi_n|$. Then (c) is a subspace of l^∞. Show that it is complete. Let u_0 be the x for which $\xi_n = 1$ for every n, and let $u_1, u_2, \cdots$ be as in the discussion of l^p. Show that

$$x = \xi_0 u_0 + \sum_1^\infty (\xi_k - \xi_0)u_k.$$

If $x' \in (c)'$ and $x'(u_k) = \alpha_k$, show by special choice of x that $\sum_1^\infty |\alpha_k| < \infty$. Then in general we can write

$$x'(x) = \left(\alpha_0 - \sum_1^\infty \alpha_k\right)\xi_0 + \sum_1^\infty \alpha_k \xi_k.$$

Show that $(c)'$ is congruent to l^1 under the correspondence $x' \leftrightarrow b = \{\beta_n\}$, where

$$x'(x) = \sum_1^\infty \beta_n \xi_{n-1}.$$

2. Let (c_0) be the closed subspace of (c) consisting of those $x \in (c)$ for which $\xi_n \to 0$. Show that $(c_0)'$ is congruent to l^1 under the correspondence $x' \leftrightarrow a = \{\alpha_n\}$, where $x'(x) = \sum_1^\infty \alpha_k \xi_k$.

3. If $1 < p < \infty$, then l^p is norm reflexive.

4. l^1 is not norm reflexive.

5. (c) and (c_0) are not norm reflexive.

6. State and prove a representation theorem for $l^p(n)$, $1 \leqslant p \leqslant \infty$.

7. Let (a, b) be a finite or infinite interval of the real axis. The general representation of a continuous linear functional on $L^p(a, b)$, if $1 \leqslant p < \infty$, is

$$x'(x) = \int_a^b v(t)x(t)\,dt,$$

where $v \in L^{p'}(a, b)$. The correspondence $x' \leftrightarrow v$ is a congruence of $(L^p)'$ and $L^{p'}$. See Banach, **1**, pages 61–65 and Riesz and Nagy, **1**, page 78. A more general form of this representation theorem is given in § 7.4.

4.4 The Principle of Uniform Boundedness

In this section we have several theorems, in each of which the conclusion is that a certain set in a certain normed linear space is bounded. In some cases the space is a space of continuous linear operators. It is convenient to state several theorems, even though, in a certain sense, they are all derivable from one general theorem. For the purpose of applications it is better to state the results in various forms adapted to easy recognition in the various situations where they are needed. But there is a single fundamental principle underlying all the theorems. This principle goes back to the fact that a complete metric space is of the second category in itself. Our proofs use this fact through the medium of the closed-graph theorem (Theorem 4.2–I).

To avoid trivial situations we assume throughout that we deal with linear spaces which contain some nonzero elements.

Theorem 4.4–A. *Let X be a normed linear space, and let F be a nonempty subset of X. Suppose that $\sup_{x\in F} |x'(x)| < \infty$ for each $x' \in X'$. Then $\sup_{x\in F} \|x\| < \infty$ (i.e., F is a bounded set).*

This theorem is merely a special case of the next theorem, but we state Theorem 4.4–A separately to make the nature of the principle of uniform boundedness stand out clearly before we embark upon more general theorems.

Theorem 4.4–A can be given a geometric interpretation by using the notion of a hyperplane. A closed hyperplane through 0 is a set of the form $\{x : x'(x) = 0\}$, where x' is some element of X'. The set $\{x : |x'(x)| < \alpha\}$ (where $\alpha > 0$) is then the set of all points between the two closed hyperplanes $\{x : x'(x) = \alpha\}$, $\{x : x'(x) = -\alpha\}$, which are parallel to one through the origin, and at equal distances from it on opposite sides of 0. Theorem 4.4–A states that, if for every closed hyperplane L through 0 the set F lies in some open "slab" between two closed hyperplanes parallel to L, then F is bounded.

To generalize Theorem 4.4–A we propose to draw the same conclusion as before but with the weaker assumption that $\sup_{x\in F} |x'(x)| < \infty$ for each $x' \in M$, where M is merely a subset of X' but of a suitably special nature.

Definition. If M is a linear manifold in X' and $x \in X$, let

$$p_M(x) = \sup |x'(x)| \qquad \text{for } x' \in M \text{ and } \|x'\| = 1. \tag{4.4–1}$$

Let

$$\nu(M) = \inf p_M(x) \qquad \text{for } x \in X \text{ and } \|x\| = 1. \tag{4.4–2}$$

We call $\nu(M)$ the *characteristic* of M. If $\nu(M) > 0$, M is said to be *norm-determining* for X.

It is evident that

$$(4.4\text{–}3) \qquad \nu(M)\|x\| \leqslant p_M(x) \leqslant \|x\|$$

and that $0 \leqslant \nu(M) \leqslant 1$. If $\nu(M) > 0$, then p_M is evidently a norm on X, and the topology for X defined by using this norm is the same as the topology for X defined by using $\|x\|$ as norm.

Theorem 4.4–B. *Let F be a nonempty subset of the normed linear space X, and let M be a closed linear manifold in X' which is norm-determining for X. Suppose that $\sup_{x \in F} |x'(x)| < \infty$ for each $x' \in M$. Then F is a bounded set.*

PROOF. Consider $B(F)$, the Banach space of all bounded scalar-valued functions f defined on F, with $\|f\| = \sup_{x \in F} |f(x)|$. Define a linear operator T with domain M and range in $B(F)$ by defining Tx' to be the element of $B(F)$ whose value at x is $x'(x)$. The hypothesis assures that $Tx' \in B(F)$, and T is obviously linear. Since $\|x_n' - x'\| \to 0$ implies $x_n'(x) \to x'(x)$ for each $x \in X$, it is a simple matter (using Theorem 4.2–A) to verify that T is a closed operator. But, since M and $B(F)$ are both complete, it follows by the closed-graph theorem that T is continuous. Hence $|x'(x)| \leqslant \|T\| \, \|x'\|$ if $x \in F$ and $x' \in M$. It follows that $\nu(M)\|x\| \leqslant p_M(x) \leqslant \|T\|$ if $x \in F$. Since $\nu(M) > 0$, the fact that F is bounded now follows at once.

If $M = X'$, then $p_M(x) = \|x\|$, by (4.3–2). Hence $\nu(M) = 1$ in this case. It is then clear that Theorem 4.4–A is just a special case of Theorem 4.4–B.

Next we consider X' and X''. If J is the operator which maps X canonically into X'', the range of J is a subspace of X'' which is norm-determining for X'. In fact, the characteristic of $\mathscr{R}(J)$ is 1, as we see by examining the definition of $\nu\{\mathscr{R}(J)\}$ and recalling the definition of $\|x'\|$. It is furthermore true that, if X is complete, $\mathscr{R}(J)$ is closed in X''. We can therefore apply Theorem 4.4–B to subsets of X', using $\mathscr{R}(J)$ for the role of M. This gives us the result:

Theorem 4.4–C. *Let X be a Banach space, and let F be a subset of X'. Suppose that $\sup_{x' \in F} |x'(x)| < \infty$ for each $x \in X$. Then $\sup_{x' \in F} \|x'\| < \infty$.*

As immediate application of Theorems 4.4–A and 4.4–C we have:

Theorem 4.4–D. (*a*) *Let $\{x_n\}$ be a sequence in the normed linear space X such that $\lim_{n \to \infty} x'(x_n)$ exists for each $x' \in X$. Then the sequence of norms $\|x_n\|$ is bounded.*

(*b*) *Let X be a Banach space, and let $\{x_n'\}$ be a sequence in X'* such *that* $\lim_{n\to\infty} x_n'(x)$ *exists for each $x \in X$. Then, the sequence of norms $\|x_n'\|$ is bounded.*

We can formulate a theorem similar to Theorem 4.4–B, dealing with uniform boundedness of a set of bounded linear operators. Let X and Y be normed linear spaces, and let Z be a subspace of X. If $T \in [X, Y]$, define

$$q_Z(T) = \sup \|Tx\| \qquad \text{for } x \in Z \text{ and } \|x\| = 1.$$

Then define

$$\mu(Z) = \inf q_Z(T) \qquad \text{for } T \in [X, Y] \text{ and } \|T\| = 1.$$

If $\mu(Z) > 0$, we shall say that Z is norm-determining for $[X, Y]$. [Note that $\mu(X) = 1$.] Evidently $0 \leqslant \mu(Z) \leqslant 1$ and

$$\mu(Z)\|T\| \leqslant q_Z(T) \leqslant \|T\|.$$

We have the following result:

Theorem 4.4–E. *Suppose X is a Banach space, and let Y be a normed linear space. Let Z be a closed subspace of X which is norm-determining for $[X, Y]$. Let G be a nonempty subset of $[X, Y]$ such that* $\sup_{T\in G} \|Tx\| < \infty$ *for each $x \in Z$. Then* $\sup_{T\in G} \|T\| < \infty$.

PROOF. The argument is essentially similar to that in the proof of Theorem 4.4–B. First of all, we observe that it suffices to prove the theorem with the added assumption that Y is complete. For, if Y is not complete and $\hat{Y}$ is its completion, we can just as well look upon G as a subset of $[X, \hat{Y}]$. This does not affect $\|Tx\|$ or $\|T\|$. Now, assuming that Y is complete, we consider the class of all bounded functions g with domain G and range in Y, with $\|g\| = \sup_{T\in G} \|g(T)\|$. This space is complete. For each $x \in Z$ let Ax be the function on G with value Tx at T. Then A is a linear operator with domain Z. It is easily proved that A is closed and hence continuous. The proof that G is a bounded set in $[X, Y]$ is then just like the last part of the proof of Theorem 4.4–B.

It is possible to regard Theorem 4.4–B as a special case of Theorem 4.4–E. The parts of X, Y, $[X, Y]$, Z, and G in Theorem 4.4–E are played by X', A (the space of scalars), X'', M and $J(F)$ in Theorem 4.4–B. Saying that $J(F)$ is bounded in X'' is equivalent to saying that F is bounded in X (J the canonical mapping).

As a particular case of Theorem 4.4–E we note the following (with assumptions on X and Y as in the theorem): *If* $T_n \in [X, Y]$ *and if* $Tx = \lim_{n\to\infty} T_n x$ *exists for each* $x \in X$, *then* $T \in [X, Y]$; *that is,* T *is not only linear, but also continuous.*

One of the striking applications of the principle of uniform boundedness is concerned with the notion of analyticity for vector-valued functions of a complex variable. If X is a complex Banach space and Δ is an open set in the complex plane, we shall say that a function f defined on Δ, with values in X, is *locally analytic* on Δ if $f(\lambda)$ is differentiable with respect to the complex variable λ at each point of Δ. If $f'(\lambda_0)$ is the derivative of $f(\lambda)$ at λ_0, this means that

$$\left\| \frac{f(\lambda) - f(\lambda_0)}{\lambda - \lambda_0} - f'(\lambda_0) \right\| \to 0 \qquad \text{as } \lambda \to \lambda_0.$$

A necessary condition for this to be true is that

$$\lim_{\lambda\to\lambda_0} x'\left\{\frac{f(\lambda) - f(\lambda_0)}{\lambda - \lambda_0}\right\}$$

exist for each $x' \in X'$ (the limit, of course, will be $x'\{f'(\lambda_0)\}$). The remarkable fact is that this, apparently weak, necessary condition is in fact also a sufficient condition for the local analyticity of f.

Theorem 4.4–F. *Let X be a complex Banach space, and let f be a function with values in X, defined on an open set Δ in the complex λ-plane. Let M be a closed linear manifold in X' which is norm-determining for X. Suppose that, for each $x' \in M$, $x'[f(\lambda)]$ is differentiable at each point of Δ. Then f is locally analytic on Δ.*

PROOF. Since X is complete, it will suffice to prove that for each point λ_0 in Δ the expression

$$\frac{f(\lambda) - f(\lambda_0)}{\lambda - \lambda_0} - \frac{f(\mu) - f(\lambda_0)}{\mu - \lambda_0}$$

approaches 0 as λ and μ independently approach λ_0. Let $r > 0$ be such that $\lambda \in \Delta$ if $|\lambda - \lambda_0| \leqslant r$. Suppose $0 < |\lambda - \lambda_0| < r$ and $0 < |\mu - \lambda_0| < r$. By Cauchy's formula we have, for each $x' \in M$,

$$x'(f(\lambda)) = \frac{1}{2\pi i}\int_C \frac{x'(f(\xi))}{\xi - \lambda}\, d\xi,$$

where C is the circle $|\xi - \lambda_0| = r$, positively oriented. Corresponding formulas hold with λ_0 and μ in place of λ. A straightforward calculation leads to the formula

$$(4.4\text{–}4)\quad x'\left\{\frac{f(\lambda)-f(\lambda_0)}{\lambda-\lambda_0}-\frac{f(\mu)-f(\lambda_0)}{\mu-\lambda_0}\right\} = \frac{1}{2\pi i}\int_C x'[f(\xi)]\frac{\lambda-\mu}{(\xi-\lambda)(\xi-\mu)(\xi-\lambda_0)}\,d\xi.$$

Now $x'[f(\xi)]$ is continuous, and hence bounded, on C. Therefore, by Theorem 4.4–B, there is some constant A such that $\|f(\xi)\| \leqslant A$ when ξ is on C. If we now require that $|\lambda-\lambda_0| \leqslant \frac{1}{2}r$, $|\mu-\lambda_0| \leqslant \frac{1}{2}r$, so that $|\xi-\lambda| \geqslant \frac{1}{2}r$ when ξ is on C (and likewise for μ in place of λ), we readily see that the absolute value of the left member of (4.4–4) does not exceed $4r^{-2}A\|x'\|\,|\lambda-\mu|$. It follows from (4.4–1) and (4.4–3) that

$$\left\|\frac{f(\lambda)-f(\lambda_0)}{\lambda-\lambda_0}-\frac{f(\mu)-f(\lambda_0)}{\mu-\lambda_0}\right\| \leqslant \frac{4A|\lambda-\mu|}{r^2v(M)}$$

if $0<|\lambda-\lambda_0| \leqslant \frac{1}{2}r$ and $0<|\mu-\lambda_0| \leqslant \frac{1}{2}r$. The desired conclusion now follows.

In order to emphasize the remarkable quality of Theorem 4.4–F, we point out explicitly that, if $\{x_n\}$ is a sequence in X such that $x'(x_n) \to x'(x)$ for each $x' \in X'$, *it does not necessarily follow* that $\|x_n - x\| \to 0$. For example, take $X = l^2$ and $x_1 = (1, 0, 0, \cdots)$, $x_2 = (0, 1, 0, \cdots)$, $x_3 = (0, 0, 1, 0, \cdots)$, etc. Then $x'(x_n) \to 0$ for each $x' \in X'$. For, we know by Theorem 4.32–A that x' is representable in the form $x'(x) = \sum_1^\infty \alpha_k\xi_k$, with $\sum_1^\infty |\alpha_k|^2 < \infty$. Hence $x'(x_n) = \alpha_n \to 0 = x'(0)$ for each x'. But it is not true that $\|x_n\| \to 0$.

There is a theorem which is related to Theorem 4.4–F in much the same way that Theorem 4.4–E is related to Theorem 4.4–B.

Theorem 4.4–G. *Let X and Y be complex Banach spaces, and let Δ be an open set in the complex plane. For each λ in Δ let A_λ be an element of $[X, Y]$. Suppose that $A_\lambda x$ is a locally analytic function on Δ, with values in Y, for each x in X. Then A_λ is a locally analytic function on Δ, with values in $[X, Y]$.*

We have here an inference from differentiability in the metric of Y to differentiability in the metric of $[X, Y]$. Or, to put it another way, if $A_\lambda x$ is differentiable for each x, it is differentiable uniformly for all x such that $\|x\| \leqslant 1$. As in Theorem 4.4–E, it is sufficient to make the assumption for all x in a closed subspace of X which is norm-determining for $[X, Y]$. The proof is left to the reader.

PROBLEMS

1. It is essential that X be complete in Theorem 4.4–C. To show this let X be the subset of those $x \in l^2$ for which $x = \{\xi_k\}$ and $\xi_k = 0$ if k exceeds some integer depending on x. Let $x_n'(x) = n\xi_n$, and let F be the countable set of the elements $x_1', x_2', \cdots$.

2. *a.* A Banach space X is said to have a *countable basis* $\{u_n\}(n = 1, 2, \cdots)$ if each $x \in X$ can be represented in one and only one way as a series $x = \sum_1^\infty \omega_k u_k$. Show that no u_n is 0 and that $\{u_n/\|u_n\|\}$ is also a countable basis.

b. Let W be the class of all sequences $w = \{\omega_n\}$ such that $\sum_1^\infty \omega_k u_k$ is convergent. Define

$$\|w\| = \sup_n \left\| \sum_1^n \omega_k u_k \right\|.$$

Show that with this norm W is a Banach space. [First show that $|\omega_n| \leqslant 2(\|w\|/\|u_n\|)$. Now define $Tw = \sum_1^\infty \omega_k u_k$. Show that T is a linear homeomorphism of W onto all of X.

c. Define $u_k' \in X'$ by $u_k'(x) = \omega_k$. Show that $\|u_k'\| \leqslant 2\|T^{-1}\|/\|u_k\|$.

d. Let Y be the space of all scalar sequences $y = \{\eta_k\}$ such that $\sum_1^\infty \omega_k \eta_k$ is convergent whenever $w = \{\omega_k\} \in W$. Define

$$\|y\| = \sup_{\|x\| \leqslant 1} \left| \sum_1^\infty \omega_k \eta_k \right|, \text{ where } x = Tw.$$

Show that Y is congruent to X', with $y \leftrightarrow x'$, where $x'(x) = \sum_1^\infty \omega_k \eta_k$.

3. Show that a Banach space with a countable basis is separable.

4. There is an alternative way of proving Theorem 4.4–B without using the closed-graph theorem. For each $x \in F$ consider $|x'(x)|$ as a continuous, real-valued function of x', defined on M. Since M is complete we can apply Theorem 2.41–D to conclude that $|x'(x)|$ is uniformly bounded as x varies over F and x' varies over some closed sphere in M. Suppose $K > 0$, $\delta > 0$, and that $|x'(x)| \leqslant K$ if $x \in F$ and $\|x' - x_0'\| \leqslant \delta$ (x' and $x_0' \in M$). Show that $p_M(x) \leqslant 2K/\delta$ if $x \in F$ and hence that $\|x\| \leqslant 2K/\delta\nu(M)$ if $x \in F$.

5. Let $X = l^p$, $1 \leqslant p < \infty$. Let $\{\phi_n\}$ be a sequence of complex-valued functions defined on the open set Δ in the complex plane. Then $f(\lambda) = \{\phi_n(\lambda)\}$ defines a function on Δ into X which is locally analytic on Δ if and only if each ϕ_n is locally analytic and the series $\sum_1^\infty |\phi_n(\lambda)|^p$ is convergent and bounded on each compact subset of Δ. Use Theorem 4.32–A.

6. Let S be a nonempty subset of the normed linear space X, and let M be a subspace of X'. Consider the topology $\mathscr{T}_0$ for X defined by the norm and the topology $\mathscr{T}(X, M)$ as defined in § 3.81. If S is bounded relative to $\mathscr{T}_0$, it is bounded relative to $\mathscr{T}(X, M)$. Show that the converse statement is true if M is norm-determining for X.

7. Let p_1 and p_2 be two norms on the linear space X, generating topologies $\mathscr{T}_1$ and $\mathscr{T}_2$ respectively. Let M_i be the subspace of all $x' \in X^f$ such that x' is continuous relative to $\mathscr{T}_i (i = 1, 2)$. Now $\mathscr{T}_2$ is stronger than $\mathscr{T}_1$ if and only if there is some $C > 0$ such that $p_1(x) \leqslant Cp_2(x)$ for all x. Show that this is equivalent to having $M_1 \subset M_2$. As a result, $\mathscr{T}_1 = \mathscr{T}_2$ if and only if $M_1 = M_2$.

8. Let X be a locally convex topological linear T_1-space, with topology $\mathscr{T}$. Let M be the class of all $x' \in X^f$ which are continuous relative to $\mathscr{T}$. Suppose there is in $\mathscr{T}$ a neighborhood of 0, say V_0, which we may as well assume is balanced and convex, which is bounded relative to $\mathscr{T}(X, M)$. Then $\mathscr{T}$ is the topology generated by the Minkowski functional of V_0 (call it p_0), this functional being in fact a norm (see problem 2, § 3.81). First observe the following: if p is any continuous seminorm on X, let $q(x) = \max \{p(x), p_0(x)\}$, and let Q be the family of seminorms q obtained in this way. Then Q generates $\mathscr{T}$, and each q is a norm. Let $\mathscr{T}_q$ be the topology generated by the norm q, and let M_q be the set of those $x' \in X^f$ which are continuous relative to $\mathscr{T}_q$. Then $M_q \subset M$. As a result, V_0 is bounded relative to $\mathscr{T}_q$ (see problem 6). It can then be shown that p_0 and q generate the same topology, which must be $\mathscr{T}$.

9. Let M be a norm-determining linear manifold in X', with characteristic $\nu = \nu(M)$. Let $S_\alpha = \{x : \|x\| \leqslant \alpha\}$. Show that, relative to the topology $\mathscr{T}(X, M)$, the closure of S_1 lies in $S_{1/\nu}$. Hence, if $\nu = 1$, S_1 is closed in this topology. For further study of $\nu(M)$ and $\mathscr{T}(X, M)$ see Dixmier, 1.

10. If J is the canonical mapping of X into X'', $J(X)$ is norm-determining for X' and of characteristic 1. Hence $\{x' : \|x'\| \leqslant 1\}$ is closed in X' with respect to $\mathscr{T}(X', J(X))$.

11. The space X is norm reflexive if and only if the topologies $\mathscr{T}(X', X'')$ and $\mathscr{T}(X', J(X))$ for X' are the same. (A clue will be found in § 3.81.)

4.41 Weak Convergence

Suppose X is a normed linear space, and consider the weak topology $\mathscr{T}(X, M)$ for X, as defined in § 3.81, on the supposition that M is a total subspace of X'. If $\{x_n\}$ is a sequence in X, we shall say that $\{x_n\}$ is M-weakly convergent to x (where $x \in X$) if $\{x_n\}$ converges to x in the sense of the topology $\mathscr{T}(X, M)$. In view of the definition of $\mathscr{T}(X, M)$, this means the same as saying that $x'(x_n) \to x'(x)$ for each $x' \in M$. The requirement that M be total makes the limit x unique, for then $\mathscr{T}(X, M)$ is a Hausdorff topology (Theorems 3.81–B and 3.3–G). Theorem 4.4–B shows that an M-weakly convergent sequence is bounded in norm if M is a closed subspace of X' which is norm-determining for X.

In the case of X'-weak convergence (i.e., when $M = X'$) it is the common practice to drop the reference to X' and speak merely of weak convergence.

If J is the canonical mapping of X into X'', then $J(X)$ is a total subset of X''. We shall habitually write $\mathscr{T}(X', X)$ in place of $\mathscr{T}(X', J(X))$. See the remarks about duality in § 3.81; we have here an instance of this kind of duality when we consider the topology $\mathscr{T}(X, X')$ for X and the topology $\mathscr{T}(X', X)$ for X'. The topology $\mathscr{T}(X', X)$ is what is often called the w^* (or weak-*) topology for X' in current literature.

If $\{x_n'\}$ is a sequence in X' which is convergent to x' (where $x' \in X'$) in the sense of the topology $\mathscr{T}(X', X)$, we say that $\{x_n'\}$ is X-weakly convergent to x'. This happens if and only if $x_n'(x) \to x'(x)$ for each $x \in X$. When X is complete this implies that $\{\|x_n'\|\}$ is bounded [Theorem 4.4–D(*b*)].

Theorem 4.41–A. *If X is separable, every bounded sequence in X' contains an X-weakly convergent subsequence.*

PROOF. Let $\{x_n'\}$ be a bounded sequence in X', and let $\{x_k\}$ be a sequence which is dense in X. Since $\{x_n'(x_1)\}$ is a bounded sequence of scalars, it contains a bounded subsequence, which we denote by $\{x_{n1}'(x_1)\}$. Likewise $\{x_{n1}'(x_2)\}$ contains a convergent subsequence, which we denote by $\{x_{n2}'(x_2)\}$. Continuing by induction, we obtain a "diagonal sequence" $\{x_{nn}'\}$ such that $\lim\limits_{n\to\infty} x_{nn}'(x_k)$ exists for each k. It is then easy to see that $\{x_{nn}'(x)\}$ is a Cauchy sequence for each x, thus defining $x' \in X'$ such that $\{x_{nn}'\}$ is X-weakly convergent to x'.

If we suppose that $\{x_n\}$ is X'-weakly convergent to x, it follows that $\|x\| \leqslant \liminf\limits_{n\to\infty} \|x_n\|$. For, $|x'(x_n)| \leqslant \|x'\|\,\|x_n\|$, whence $|x'(x)| \leqslant \|x'\| \liminf\limits_{n\to\infty} \|x_n\|$. The desired result now follows from (4.3–2). In particular, if $\|x_n\| \leqslant C$, it follows that $\|x\| \leqslant C$ also.

Theorem 4.41–B. *If X is norm reflexive, each bounded sequence from X contains an X'-weakly convergent subsequence. In particular, if $\{x_n\}$ is a sequence for which $\|x_n\| \leqslant 1$, it contains a subsequence converging X'-weakly to a limit x for which $\|x\| \leqslant 1$.*

PROOF. In view of the remarks preceding the theorem, it suffices to prove the first assertion. Suppose $x_n \in X$ and $\sup\limits_n \|x_n\| = C < \infty$. Let X_0 be the closed linear manifold generated by $x_1, x_2, \cdots$. It is easy to see that X_0 is separable. It is reflexive, also (Theorem 4.31–B). Hence X_0'' is separable, and therefore X_0' is separable (Theorem 4.3–E). The canonical mapping of X_0 onto X_0'' carries $\{x_n\}$ into a bounded sequence in X_0''. By Theorem 4.41–A this latter sequence contains an X_0'-weakly convergent subsequence. This subsequence corresponds to a subsequence of $\{x_n\}$ which is X_0'-weakly convergent to a limit in X_0. Since every element of X', when restricted to X_0, determines an element of X_0', it is clear that the subsequence in question is X'-weakly convergent. This completes the proof.

Weak convergence is used in various contexts in the calculus of variations and in the general theory of differential equations.

PROBLEMS

1. If M is a total subspace of X', $\{x_n\}$ is M-weakly convergent to x provided that $\sup\limits_n \|x_n\| < \infty$ and that $x'(x_n) \to x'(x)$ for each $x' \in G$, where G is some dense subset of M.

2. If $\{x_n'\}$ is X-weakly convergent to x', then $\|x'\| \leqslant \liminf\limits_{n\to\infty} \|x_n'\|$.

3. If $x_n = \{\xi_k^{(n)}\} \in l^p$, where $1 < p < \infty$, $\{x_n\}$ is X'-weakly convergent to $x = \{\xi_k\}$ (where $X = l^p$) if and only if $\sup\limits_n \|x_n\| < \infty$ and $\lim\limits_{n\to\infty} \xi_k^{(n)} = \xi_k$ for each k.

4. In $X = l^1$, $\{x_n\}$ converges X'-weakly to x if and only if $\|x_n - x\| \to 0$ (see Banach **1**, page 137). In spite of this, the topology $\mathscr{T}(X, X')$ is not the same as the topology generated by the norm.

5. If $X = C[a, b]$, a sequence $\{x_n\}$ is X'-weakly convergent to x if and only if $\sup\limits_n \|x_n\| < \infty$ and $x_n(t) \to x(t)$ for each t. The necessity of these conditions is evident. The sufficiency is clear from standard convergence theorems as soon as one knows how to represent elements of X' with the aid of a suitable signed measure. See § 7.7; also see Banach, **1**, page 134.

6. For $X = L^p(a, b)$, where $1 < p < \infty$, $\{x_n\}$ converges X'-weakly to x if and only if $\sup\limits_n \|x_n\| < \infty$ and $\int_E x_n(t)\,dt \to \int_E x(t)\,dt$ for each measurable set E of finite measure on (a, b). (Use problem 1 and the known representation of elements of X' in this case.) For $p = 1$ the result breaks down, the conditions being necessary but not sufficient for weak convergence. This is because the finite linear combinations of characteristic functions of sets of finite measure are not dense in $L^\infty(a, b)$. For a further discussion see Banach **1**, page 136.

7. A space X is called X'-weakly sequentially complete if the existence of $\lim\limits_{n\to\infty} x'(x_n)$ for each $x' \in X'$ implies the existence of $x \in X$ such that $\{x_n\}$ converges X'-weakly to x. The space $C[a, b]$ is not X'-weakly sequentially complete. However, any norm-reflexive space is X'-weakly sequentially complete. The nonreflexive spaces $L^1(a, b)$ and l^1 are X'-weakly sequentially complete. See Banach **1**, page 141.

4.42 An Application of Vector-Valued Analytic Functions

Many features of the classical theory of analytic functions are retained in the theory of analytic functions whose values lie in a complex Banach space. In Chapter 5 we use the theorems of Cauchy, Liouville, and Laurent in this more general setting. In this section we present an interesting application of the theory of Banach-space-valued analytic functions to prove some classical theorems about mean values of complex-valued analytic functions.

Analyticity of a vector-valued function may be defined in terms of differentiability, as in the latter part of § 4.4. We may also use the theory of power series in a complex variable, with Banach space elements as coefficients.

We begin by observing the generalization of certain standard theorems. Here X denotes a complex Banach space.

Theorem 4.42–A. *Suppose Δ is a connected open set in the complex λ-plane. Let F be holomorphic (single valued and analytic) on Δ, with values in X, and suppose that $\|F(\lambda)\|$ is not constant on Δ. Then $\|F(\lambda)\|$ cannot attain an absolute maximum at any point of Δ.*

PROOF. We could imitate one of the classical proofs which uses Cauchy's integral formula. Instead, we shall give a proof using linear functionals. Suppose the theorem false, so that for some $\lambda_0 \in \Delta$ we have $\|F(\lambda)\| \leqslant \|F(\lambda_0)\|$ for each λ. Using Theorem 4.3–B, we choose $x' \in X'$ so that $\|x'\| = 1$ and $x'[F(\lambda_0)] = \|F(\lambda_0)\|$. Then $x'[F(\lambda)]$ is a complex-valued holomorphic function such that $|x'[F(\lambda)]| \leqslant |x'[F(\lambda_0)]|$ at all points $\lambda \in \Delta$. The classical counterpart of Theorem 4.42–A then implies that $x'[F(\lambda)]$ is constant in Δ, its value being $\|F(\lambda_0)\|$. But $|x'[F(\lambda)]| \leqslant \|F(\lambda)\|$, and, since $\|F(\lambda)\| < \|F(\lambda_0)\|$ at some points of Δ, we have a contradiction.

Now consider the special case in which Δ is the circular region $\{\lambda : |\lambda| < R\}$, where $R > 0$. Let $M(r) = \max\limits_{|\lambda| = r} \|F(\lambda)\|$, $0 \leqslant r < R$. Theorem 4.42–A implies the following theorem.

Theorem 4.42–B. *If F is holomorphic when $|\lambda| < R$, then $M(r)$ is a nondecreasing function of r. Furthermore, $0 < r_1 < r_2$ and $M(r_1) = M(r_2)$ imply that $M(r)$ is constant when $0 \leqslant r \leqslant r_2$.*

We also have the generalization of Hadamard's three-circles theorem:

Theorem 4.42–C. *If F is holomorphic and not identically 0 when $|\lambda| < R$, then $\log M(r)$ is a convex function of $\log r$ when $0 < r < R$.*

This theorem may be proved exactly as in one of the standard proofs for the complex-valued case (e.g., Titchmarsh **1**, § 5.3 and § 5.32), except that norms are written instead of absolute values.

Let us now turn to the definition of the means $\mathfrak{M}_p[f; r]$ in (3.11–9), where f is a complex-valued function which is holomorphic in the unit circle of the complex plane. We are going to assume that $1 \leqslant p < \infty$ and prove:

1. $\mathfrak{M}_p[f; r]$ *is, for* $0 \leqslant r < 1$, *a strictly increasing function of* r, *unless* f *is a constant.*

2. *If* f *is not identically* 0, $\log \mathfrak{M}_p[f; r]$ *is a convex function of* $\log r$ *when* $0 < r < 1$.

These propositions go back to G. H. Hardy, **1**. Very beautiful proofs of the propositions were given by F. Riesz, **3**, using the principle of subharmonic functions. The propositions are true for all $p > 0$, but our proofs require $p \geqslant 1$, because we depend on the fact that H^p is a Banach space if $p \geqslant 1$. See § 3.11 and problem 1*c* in § 3.13.

If f is holomorphic when $|z| < 1$ and if $|\lambda| < 1$, let f_λ be the function defined by $f_\lambda(z) = f(\lambda z)$. Then f_λ is holomorphic in a domain which includes the closed disk $|z| \leqslant 1$, so that $f_\lambda \in H^p$. Hence the function F defined by $F(\lambda) = f_\lambda$ is a function of λ with values in H^p. It is a fact of crucial importance for us that F is holomorphic when $|\lambda| < 1$. See problems 1–3 for a discussion of this matter. If we compute $\|F(\lambda)\| = \|f_\lambda\|$, using (3.11–11), we find that $\|F(\lambda)\| = \|F(|\lambda|)\|$. Thus $\|F(\lambda)\|$ is constant on the circle $|\lambda| = r$; we denote it by $M(r)$. Theorems 4.42–B and 4.42–C apply to $M(r)$ for $r < 1$. The computation of $\|F(\lambda)\|$ also shows that

$$M(r) = \sup_{0 \leqslant x < 1} \mathfrak{M}_p[f; rx]. \tag{4.42–1}$$

From this we shall be able to show that

$$M(r) = \mathfrak{M}_p[f; r]. \tag{4.42–2}$$

First we take note of the fact that

$$\mathfrak{M}_p[f; 0] = |f(0)| \leqslant \mathfrak{M}_p[f; r] \tag{4.42–3}$$

if $0 < r < 1$ and that the inequality here is strict unless f is a constant function. For, by Cauchy's formula,

$$f(0) = \frac{1}{2\pi} \int_0^{2\pi} f(re^{i\theta})\, d\theta,$$

whence $|f(0)| \leqslant \mathfrak{M}_1[f; \theta]$. The inequality here is easily seen to be strict unless f is constant. But then $\mathfrak{M}_1[f; r] \leqslant \mathfrak{M}_p[f; r]$ if $1 < p$ (see Hardy, Littlewood and Polya, **1**, page 143), whence (4.42–3) follows.

Now let $\phi(r) = \mathfrak{M}_p[f; r]$; we know that ϕ is continuous. Evidently $M(0) = \phi(0)$ and $\phi(r) \leqslant M(r)$, because of (4.42–1). Suppose that for some r_0 we have $0 < r_0 < 1$ and $\phi(r_0) < M(r_0)$. Consider the maximum of $\phi(r)$ for $0 \leqslant r \leqslant r_0$. It occurs for some r_1 on the interval and is equal to $M(r_0)$, by (4.42–1). That is, $\phi(r_1) = M(r_0)$. Hence $r_1 < r_0$, because $\phi(r_1) \neq \phi(r_0)$. We see that $M(r_0) \leqslant M(r_1)$. But $M(r_1) \leqslant M(r_0)$, by the first conclusion in Theorem 4.42–B. Hence $M(r_1) = M(r_0)$, and it follows from the second conclusion in Theorem 4.42–B that $M(r)$ is constant on $[0, r_0]$. Then $\phi(r_0) < M(r_0) = M(0) = \phi(0)$. This contradicts (4.42–3). Hence (4.42–2) is established.

Proposition 2 now follows at once from (4.42–2) and Theorem 4.42–C. From Theorem 4.42–B we infer that $\mathfrak{M}_p[f; r]$ is a nondecreasing function of r. Actually, $\mathfrak{M}_p[f; r]$ must be strictly increasing, unless f is constant, for otherwise, by Theorem 4.42–B we would have $M(r)$ constant on some interval $[0, r_1]$, whence $\mathfrak{M}_p[f; 0] = \mathfrak{M}_p[f; r_1]$, in contradiction to the statement made in connection with (4.42–3).

PROBLEMS

1. Let $u_n(z) = z^n$, $n = 0, 1, \cdots$. Show that u_n, as an element of H^p, has $\|u_n\| = 1$.

2. If $h \in H^p$ and $h(z) = \sum\limits_0^\infty c_n z^n$, use Cauchy's formulas for c_n to show that $r^n|c_n| \leqslant \mathfrak{M}_p[h; r]$ if $0 < r < 1$, and hence $|c_n| \leqslant \|h\|$. Deduce that $|h(z)| \leqslant \|h\|\,(1 - |z|)^{-1}$ if $|z| < 1$.

3. If $f(z) = \sum\limits_0^\infty a_n z^n$ converges when $|z| < 1$, show that $\sum\limits_0^\infty a_n \lambda^n u_n$ is convergent to f_λ in the metric of H^p (with u_n as in problem 1). First show that the series converges to some element $g \in H^p$. Then use problem 2 to show that $\sum\limits_0^\infty a_n \lambda^n z^n = g(z)$, whence $g = f_\lambda$. The conclusion is that $F(\lambda)$ is given by a power series in λ with coefficients from H^p. Hence F is holomorphic. For this and more general results of the same type see Taylor **3**, **4**.

4.5 The Conjugate of a Bounded Linear Operator

The notion of the conjugate of a bounded linear operator is closely related to the notion of the transpose of a linear operator, as defined in § 1.8.

Let X and Y be normed linear spaces, and suppose $A \in [X, Y]$. If $y' \in Y'$, the linear functional x' defined on X by $x'(x) = y'(Ax)$ is clearly continuous. We write $x' = A'y'$. The operator A', which is linear and maps Y' into X', is called the *conjugate* of A. The definition of A' is expressed by the formula

$$\langle x, A'y'\rangle = \langle Ax, y'\rangle, \qquad x \in X, y \in Y'. \tag{4.5–1}$$

If $Y' = Y^f$ (which is the case if Y is finite dimensional), A' is the same thing as the transpose A^T. However, Y' is a proper subspace of Y^f in general, so A^T is an extension of A', or A' is the restriction of A^T to the space Y'.

If $A'y' = x'$, we have $|x'(x)| = |y'(Ax)| \leqslant \|y'\|\,\|Ax\| \leqslant \|y'\|\,\|A\|\,\|x\|$, and so $\|x'\| \leqslant \|A\|\,\|y'\|$. This shows that A' is continuous on Y' and that $\|A'\| \leqslant \|A\|$. On the other hand,

$$|y'(Ax)| = |x'(x)| \leqslant \|x'\|\,\|x\| = \|A'y'\|\,\|x\|.$$

Hence, by (4.3–2) and the definition of $\|A'\|$,

$$\|Ax\| = \sup_{\|y'\|\leqslant 1} |y'(Ax)| \leqslant \sup_{\|y'\|\leqslant 1} \|A'y'\|\,\|x\| = \|A'\|\,\|x\|.$$

Therefore $\|A\| \leqslant \|A'\|$. We have thus proved the formula

$$\|A\| = \|A'\|. \tag{4.5–2}$$

Other notations and terminologies have been used for A'. It has been denoted by $\bar{A}$ or A^*; it has been called the *adjoint* of A and also the *dual* of A. In this book the notation A^* and the name "adjoint" are reserved for use in connection with linear operators in Hilbert space. In that situation A^* is closely related to A', but is not the same as A'. It is troublesome to have A^* mean one thing when we deal with Hilbert space and another thing in the general theory of normed linear spaces. Hence our distinction between A^* and A'.

We sometimes need some formal algebraic rules for handling conjugate operators. If A and B are in $[X, Y]$, then

$$(A + B)' = A' + B' \qquad \text{and} \qquad (\alpha A)' = \alpha A'. \tag{4.5–3}$$

If $B \in [X, Y]$ and $A \in [Y, Z]$, then

$$(AB)' = B'A'. \tag{4.5–4}$$

If $A \in [X, Y]$ and if A^{-1} exists and belongs to $[Y, X]$, then $(A')^{-1}$ exists, belongs to $[X', Y']$, and is the same as $(A^{-1})'$. We leave it for the reader to verify these assertions.

If we are studying the operator A, a knowledge of certain facts about A' can be helpful to us. In general A' is more amenable to investigation than A^T. In studying A' we usually need representation theorems for elements of Y' and X' (as in § 4.32, for example). We also need representation theorems for bounded linear operators. For instance, suppose $X = Y = l^p$, where $1 \leqslant p < \infty$. We know from Theorem 4.32–A that X' is congruent in a natural way to $l^{p'}$. In the problems at the end of § 4.51 we shall see that, if $A \in [l^p]$, A can be represented by an infinite matrix and that A', which we identify with an element of $[l^{p'}]$, is representable by the transposed matrix. Likewise, if $X = Y = L^2(a, b)$ and if K is the Fredholm-type operator defined as $y = Kx$ by

$$y(s) =^{0} \int_a^b k(s, t)x(t)\, dt,$$

where k is an $\mathscr{L}^2$ kernel (see § 4.12), we can identify X' with $L^2(a, b)$ and K' with the Fredholm-type operator determined by the transpose kernel $k'(s, t) = k(t, s)$. This identification is a consequence of the general representation theorem for linear functionals on a Hilbert space (see § 4.81).

PROBLEMS

1. Suppose X and Y are normed linear spaces, and let A be a linear operator with domain X and range in Y. Suppose that Y' contains a closed linear manifold M which is norm-determining for Y and such that $A^T(M) \subset X'$. Then A is continuous. Apply Theorem 4.4–B to the set $\{Ax : \|x\| \leqslant 1\}$ in Y.

2. If A maps X congruently onto all of Y, then A' maps Y' congruently onto all of X'.

3. Suppose $A \in [X_1, X_2]$. Then $(A')' \equiv A'' \in [X_1'', X_2'']$. Let J_1 and J_2 be the operators defining the canonical mappings of X_1 and X_2 into X_1'' and X_2'' respectively. Then $A''J_1 = J_2A$. If we identify X_1 with $J_1(X_1)$ and X_2 with $J_2(X_2)$, this means that A'' is an extension of A. In particular, A'' coincides with A if X_1 is norm reflexive.

4. If X_1 is norm reflexive and X_1 is congruent to X_2, then X_2 also is norm reflexive.

4.51 Some Representations of Bounded Linear Operators

In this section we discuss the representation of bounded linear operators on X into Y, especially for the case in which X and Y are chosen in various ways from the spaces l^p, (c) and (c_0). The discussion is based on a theorem of very general scope about the representation of linear operators.

For the purpose of our first general theorem we make the following assumptions: X is an arbitrary Banach space; Z is an arbitrary normed linear space; S is an arbitrary nonempty set of elements of unspecified nature; Y is a Banach space which is a subspace of the linear space of all functions defined on S, with values in Z. If $s \in S$ and $y \in Y$, we call $y(s)$ the component of y at s, and we assume that $y(s)$ is a continuous function of y. A possible choice for Y is l^p $(1 \leqslant p \leqslant \infty)$, with S the set of positive integers and Z the space of scalars. Again, the space $B(S)$ of bounded scalar-valued functions on S, for an arbitrary S, is a possible choice for Y.

In what follows we refer to the space $[X, Z]$. This is X' if Z happens to be the space of scalars; it is this choice of Z that we shall use mainly in the applications of these general considerations.

Theorem 4.51–A. *Assume X, Y, Z, S as in the foregoing remarks. Then, to each $A \in [X, Y]$ there corresponds a function on S to $[X, Z]$, whose value at s we denote by $a(s)$, such that, if $y = Ax$ then $y(s) = a(s)x$ for each s. Conversely, if $a(\cdot)$ is any function with domain S and range in $[X, Z]$ such that for each $x \in X$ the formula $y(s) = a(s)x$ defines an element of Y, this dependence of y on x defines an element A of $[X, Y]$.*

PROOF. If $A \in [X, Y]$ is given and $y = Ax$, we define $a(s)$ by $a(s)x = y(s)$. Then $a(s) \in [X, Z]$ for each s. The linearity of $a(s)$ is evident, and the continuity of $a(s)x$ with respect to x follows from the fact that $y(s)$ is a continuous function of y. For the converse part of the proof, the linearity of A is evident, and its continuity will follow from Theorem 4.2–I if we prove that A is closed. Suppose that $x_n \to x$. Then $a(s)x_n \to a(s)x$, because $a(s) \in [X, Z]$. Suppose also that $Ax_n \to y$. Then $(Ax_n)(s) \to y(s)$ for each s, by the continuous dependence of components on vectors. But $(Ax_n)(s) = a(s)x_n$. Hence $a(s)x = y(s)$, so that $y = Ax$; that is, A is closed.

A situation of some especial interest is that in which the space Y is of such a nature that the norm of an element in Y is given by

$$\|y\| = \sup_{s \in S} \|y(s)\|_Z. \tag{4.51–1}$$

Here $\|y(s)\|_Z$ denotes the norm of $y(s)$ as an element of Z, s being fixed. Among the examples of spaces Y of this type, all with Z the space of scalars, we cite l^∞, (c), (c_0), $B(S)$, and, if S is a topological space, the subspace $C(S)$ of $B(S)$.

Theorem 4.51–B. *With the hypotheses as in Theorem* 4.51–A, *suppose the norm in Y is given by* (4.51–1). *Then $a(\cdot)$ is a bounded function on S to $[X, Z]$, and*

$$\|A\| = \sup_{s \in S} \|a(s)\|. \tag{4.51–2}$$

If Z is complete and if Y consists of all bounded functions on S to Z, any bounded function $a(\cdot)$ on S to $[X, Z]$ determines an operator belonging to $[X, Y]$.

The proof is left to the reader. The assumption in the last part, that Z is complete, is made to insure that Y is complete, this being needed in Theorem 4.51–A.

Let us denote by $B_Z(S)$ the Banach space of all bounded functions on S to the Banach space Z, with norm as in (4.51–1). Using this notation for arbitrary Z, we see that Theorem 5.51–B implies that $[X, B_Z(S)]$ is congruent to $B_{[X, Z]}(S)$ in a natural way. If Y is a closed subspace of $B_Z(S)$, the theorem states that $[X, Y]$ is congruent to a subspace of $B_{[X, Z]}(S)$, but the theorem does not describe this subspace. A certain description of the subspace may be inferred from Theorem 4.51–A.

We can now give representation theorems for elements of $[X, Y]$, where X is one of the spaces (c_0), (c), or $l^p(1 \leqslant p < \infty)$ and Y is one of the spaces l^∞, (c), or (c_0). For this purpose we need Theorem 4.32–A and the results stated in problems 1 and 2 at the end of § 4.32. We take Z as the space of scalars and S as the set of positive integers.

Theorem 4.51–C. *Each bounded linear operator A on (c_0) into l^∞, (c), or (c_0) determines and is determined by an infinite matrix of scalars $\alpha_{ij}(i, j = 1, 2, \cdots)$, $y = Ax$ being expressed by the equations*

$$\eta_i = \sum_{j=1}^{\infty} \alpha_{ij}\xi_j \qquad i = 1, 2, \cdots. \tag{4.51–3}$$

The norm of A is given by

$$\|A\| = \sup_i \sum_{j=1}^{\infty} |\alpha_{ij}|. \tag{4.51–4}$$

For the case of mapping into l^∞ the only restriction on the matrix (α_{ij}) is that the expression defining $\|A\|$ be finite. For the case of mapping into (c) the only additional requirement is that the limits

$$\alpha_j = \lim_{i\to\infty} \alpha_{ij} \qquad j = 1, 2, \cdots \tag{4.51–5}$$

must exist. Finally, the mapping is into (c_0) if and only if, in addition, $\alpha_1 = \alpha_2 = \cdots = 0$.

PROOF. For the case of mapping into l^∞, the theorem results immediately from problem 2, § 4.32, and Theorems 4.51–A, 4.51–B. For the case of mapping into (c) the further requirement is that $\lim_{i\to\infty} \eta_i$ exist for each choice of x in (4.51–3). An obvious special choice of x shows the necessity for the existence of the limits in (4.51–5). On the other hand, if these limits exist, it is easy to prove that

$$\lim_{i\to\infty} \eta_i = \sum_{j=1}^{\infty} \alpha_j \xi_j. \tag{4.51–6}$$

For, we know that $\sum_{j=1}^{n} |\alpha_{ij}| \leqslant \|A\|$ for all i and n. Letting first i and then n become infinite we see that $\sum_{j=1}^{\infty} |\alpha_j| \leqslant \|A\|$. Then, for each n,

$$\left|\eta_i - \sum_{1}^{\infty} \alpha_j \xi_j\right| = \left|\sum_{j=1}^{\infty} (\alpha_{ij} - \alpha_j)\xi_j\right| \leqslant$$

$$\sum_{=1}^{n} |\alpha_{ij} - \alpha_j|\,|\xi_j| + 2\|A\| \sup_{j>n} |\xi_j|.$$

From this it follows that (4.51–6) is true, so that $y \in (c)$ if $x \in (c_0)$. We leave to the reader the proof of the last assertion in the theorem.

The next theorem concerns operators $A \in [(c), Y]$, where Y is either l^∞, (c), or (c_0). It is slightly different from the foregoing theorem, because of the intervention of the limit $\xi_0 = \lim_{j\to\infty} \xi_j$ when $x = \{\xi_j\} \in (c)$.

Theorem 4.51–D. *Each bounded linear operator A on (c) into l^∞, (c), or (c_0) determines and is determined by a matrix of scalars α_{ij} ($i = 1, 2, \cdots$, $j = 0, 1, 2, \cdots$), $y = Ax$ being expressed by the equations*

$$\eta_i = \sum_{j=0}^{\infty} \alpha_{ij} \xi_j \qquad i = 1, 2, \cdots. \tag{4.51–7}$$

The norm of A is given by

$$\|A\| = \sup_i \sum_{=0}^{\infty} |\alpha_{ij}|. \tag{4.51–8}$$

For $A \in [(c), l^\infty]$ the sole condition on the matrix (α_{ij}) is that the expression defining $\|A\|$ be finite. For $A \in [(c), (c)]$ there is the additional requirement that the limit

$$\alpha = \lim_{i\to\infty} \sum_{j=0}^{\infty} \alpha_{ij} \tag{4.51–9}$$

exist and that the limits

$$\alpha_j = \lim_{i\to\infty} \alpha_{ij} \tag{4.51–10}$$

exist if $j = 1, 2, \cdots$ *(no requirement when* $j = 0$*). Finally, the range of* A *is in* (c_0) *if and only if, in addition,* $\alpha = 0$ *and* $\alpha_1 = \alpha_2 = \cdots = 0$.

Proof is left to the reader, who should refer to problem 1, § 4.32.

We note, incidentally, that, if $A \in [(c), (c)]$ and $\lim_{i\to\infty} \eta_i = \eta_0$, then

$$\eta_0 = \xi_0\alpha + \sum_{j=1}^{\infty} (\xi_j - \xi_0)\alpha_j. \tag{4.51-11}$$

For, from problem 1, § 4.32 we have

$$Ax = \xi_0 Au_0 + \sum_{j=1}^{\infty} (\xi_j - \xi_0)Au_j.$$

If u_0' is the continuous linear functional defined by $u_0'(x) = \xi_0$, we have

$$u_0'(Ax) = \xi_0 u_0'(Au_0) + \sum_{j=1}^{\infty} (\xi_j - \xi_0)u_0'(Au_j),$$

and a little checking shows that this result is equivalent to (4.51–11). We see from (4.51–11) that $\eta_0 = \xi_0$ for every choice of x if and only if $\alpha = 1$ and $\alpha_1 = \alpha_2 = \cdots = 0$.

When X and Y both are spaces with a countable basis, they may be regarded as spaces whose elements are sequences. This latter is then true of the conjugate spaces X', Y'. (For basic facts about X and X' when X has a countable basis see problem 2, § 4.4.) In such cases an element $A \in [X, Y]$ is representable by an infinite matrix, and the conjugate operator A' is then representable by the transpose of this matrix. For more precise details see problem 6.

PROBLEMS

1. Each bounded linear operator A on l^p $(1 \leqslant p < \infty)$ into l^∞, (c), or (c_0) is representable by an infinite matrix (α_{ij}) of scalars, where $y = Ax$ is expressed by equations (4.51–3). The norm of A is

$$\|A\| = \begin{cases} \sup\limits_i \left(\sum\limits_{j=1}^{\infty} |\alpha_{ij}|^{p'} \right)^{1/p'} & \text{if } 1 < p < \infty \\ \sup\limits_i \sup\limits_j |\alpha_{ij}| & \text{if } p = 1. \end{cases}$$

Except for the difference in the expression for $\|A\|$, the rest of the assertion here is just like that in Theorem 4.51–C.

2. A matrix (α_{ij}) $(i, j = 1, 2, \cdots)$ defines a bounded linear operator A on l^∞ into l^∞, by means of equations (4.51–3) if the following expression, which then defines the norm of A, is finite:

$$\|A\| = \sup_i \sum_{j=1}^{\infty} |\alpha_{ij}|.$$

This is not, however, the most general form for an element of $[l^\infty, l^\infty]$, for elements of $(l^\infty)'$ are not all expressible by an infinite series. The operator has its range in (c) or c_0) (subspaces of l^∞) under the same conditions for these things as are stated in Theorem 4.51–C.

3. If X is any Banach space, the general form of $A \in [X, l^q]$, where $1 \leqslant q < \infty$, is $y = Ax$, where $y = \{\eta_i\}$ and $\eta_i = x_1'(x)$, $\{x_i'\}$ being a sequence of elements of X' such that $\{x_i'(x)\} \in l^q$ for each x.

4. Suppose $X = l^p$ $(1 < p < \infty)$ in problem 3. The form of x_i' is $x_i'(x) = \sum_{j=1}^{\infty} \alpha_{ij}\xi_j$, where

$$\|x_i'\| = \left(\sum_{j=1}^{\infty} |\alpha_{ij}|^{p'}\right)^{1/p'}, \quad p' = \frac{p}{p-1}.$$

No direct condition on the matrix elements α_{ij} have been discovered which are necessary and sufficient for the matrix to yield an operator A in $[l^p, l^q]$. Nor is there any known explicit representation of $\|A\|$ as a function of the matrix elements. For $p = 1$ see problem 5.

5. Each bounded linear operator A on l^1 into l^q $(1 \leqslant q < \infty)$ is representable as in (4.51–3) by a matrix (α_{ij}). The only condition on the matrix is that the following expression, defining $\|A\|$, be finite.

$$\|A\| = \sup_j \left(\sum_{i=1}^{\infty} |\alpha_{ij}|^q\right)^{1/q}.$$

Conversely, any matrix satisfying this condition determines an A in $[l^1, l^q]$.

6. Let X and Y be Banach spaces with countable bases $\{u_n\}$ and $\{v_n\}$ respectively $(n = 1, 2, \cdots)$. See problem 2, § 4.4. Let $\{u_n'\}$ and $\{v_n'\}$ be the corresponding sequences of coefficient functionals, so that

$$x = \sum_1^\infty \xi_j u_j \qquad u_j'(x) = \xi_j, \qquad x \in X$$

$$y = \sum_1^\infty \eta_i v_i \qquad v_i'(y) = \eta_i, \qquad y \in Y.$$

By the problem referred to, elements of X' and Y' are representable in the respective forms

$$x'(x) = \sum_1^\infty \lambda_j \xi_j, \qquad y'(y) = \sum_1^\infty \mu_i \eta_i.$$

Suppose $A \in [X, Y]$, $\alpha_{ij} = v_i'(Au_j)$, so that $y = Ax$ is expressed by

$$\eta_i = \sum_{j=1}^{\infty} \alpha_{ij}\xi_j, \qquad i = 1, 2, \cdots.$$

Show that $x' = A'y'$ is expressed by

$$\lambda_j = \sum_{i=1}^{\infty} \alpha_{ij}\mu_i \qquad j = 1, 2, \cdots.$$

Note that this result means that, if A is represented by a certain infinite matrix, A' is represented by the transposed matrix.

7. Suppose $A \in [(c), (c)]$ is defined by (4.51–7). Let

$$\alpha_{00} = \alpha - \sum_{j=1}^{\infty} \alpha_j$$

$$\alpha_{0j} = \alpha_j, \qquad j = 1, 2, \cdots,$$

where α and α_j are defined by (4.51–9) and (4.51–10). If x' and y' are elements of $(c)'$ defined by

$$x'(x) = \sum_{j=0}^{\infty} \lambda_j\xi_j, \qquad y'(y) = \sum_{i=0}^{\infty} \mu_i\eta_i,$$

show that $Ay' = x'$ means

$$\lambda_j = \sum_{i=0}^{\infty} \alpha_{ij}\mu_i \qquad j = 0, 1, 2, \cdots.$$

4.52 The M. Riesz Convexity Theorem

Consider a matrix (α_{ij}) of complex numbers, not all 0, with m rows and n columns. Let $\xi_1, \cdots, \xi_n$ and $\zeta_1, \cdots, \zeta_m$ be complex numbers, and let $x = (\xi_1, \cdots, \xi_n)$, $z = (\zeta_1, \cdots, \zeta_m)$. If $\mu \geqslant 0$, define

$$M_\mu(x) = \begin{cases} \left(\displaystyle\sum_{j=1}^{n} |\xi_j|^{1/\mu}\right)^{\mu} & \text{if } \mu > 0 \\ \sup\limits_j |\xi_j| & \text{if } \mu = 0. \end{cases}$$

If $\nu \geqslant 0$, define $N_\nu(z)$ in a similar manner. Let

$$M(\mu, \nu) = \sup \left| \sum_{i=1}^{m} \sum_{j=1}^{n} \alpha_{ij}\zeta_i\xi_j \right|, \tag{4.52–1}$$

the supremum being taken for all x and z such that $M_\mu(x) \leqslant 1$ and $N_\nu(z) \leqslant 1$. Then $\log M(\mu, \nu)$ *is a convex function of the point* (μ, ν) *in the* *part of the* (μ, ν)*-plane defined by the inequalities* $\mu \geqslant 0$, $\nu \geqslant 0$. This means that if (μ_1, ν_1) and (μ_2, ν_2) are admissible points,

$$\log M[(1 - t)\mu_1 + t\mu_2, (1 - t)\nu_1 + t\nu_2]$$

is a convex function of t when $0 \leqslant t \leqslant 1$.

We shall refer to this result as the M. Riesz convexity theorem. M. Riesz's proof of this theorem was given in his paper **1**, subject to the more restrictive conditions $0 \leqslant \mu \leqslant 1$, $0 \leqslant \nu \leqslant 1$, $\mu + \nu \geqslant 1$. See also Hardy, Littlewood and Polya, **1**, pages 214–219. In the more general form the result is due to Thorin, **1**. If the matrix elements α_{ij} are real and if the vectors x, z are assumed to have real components, the convexity property of $\log M(\mu, \nu)$ obtains when $0 \leqslant \mu$, $0 \leqslant \nu$, and $\mu + \nu \geqslant 1$. This was shown partly by Riesz and partly by Thorin.

We shall not prove the Riesz convexity theorem. Our concern is to discuss the application of it to continuous linear mappings of l^p into l^q. For a more far-reaching discussion of the development of such ideas, with various important applications to integrals, Fourier coefficients, and other topics, see Dunford and Schwartz, **1**, §§ 10, 11 in Chapter VI, and Zygmund, **1**, Chapter XI, especially pages 189–202.

Suppose $1 \leqslant p \leqslant \infty$, $1 \leqslant q \leqslant \infty$, and consider the m by n matrix (α_{ij}) as representing a linear operator A mapping $l^q(n)$ into $l^q(m)$, with $y = Ax$ meaning $x = (\xi_1, \cdots, \xi_n)$, $y = (\eta_1, \cdots, \eta_m)$, and

$$\eta_i = \sum_{j=1}^{n} \alpha_{ij}\xi_j.$$

With the aid of Theorem 4.3–B we see that $\|A\| = \sup |y'(Ax)|$, the supremum being taken for all $x \in l^p(n)$ with $\|x\| \leqslant 1$ and all linear functionals y' on $l^q(m)$ with $\|y'\| \leqslant 1$. Since $y'(y)$ has a representation

$$y'(y) = \sum_{i=1}^{m} \zeta_i \eta_i,$$

we see that

$$\|A\| = \sup \left| \sum_{i=1}^{m} \sum_{j=1}^{n} \alpha_{ij}\zeta_i\xi_j \right|, \tag{4.52–2}$$

the supremum being taken for all x such that $\|x\| \leqslant 1$ and all $(\zeta_1, \cdots, \zeta_m)$ determining a functional y' with $\|y'\| \leqslant 1$. Now, the normed conjugate of $l^q(m)$ is congruent to $l^{q'}(m)$. Hence, comparing (4.52–1) and (4.52–2), we see that

$$\|A\| = M(1/p, 1/q'),$$

where $1/p$ is understood to be 0 if $p = \infty$, and likewise for $1/q'$. To emphasize the dependence of $\|A\|$ on p and q we shall write it as $\|A\|_{p,\,q}$.

It is clear that, when $\nu \leqslant 1$, convexity as a function of (μ, ν) is the same as convexity as a function of $(\mu, 1 - \nu)$. Hence, by Riesz's theorem, $\log \|A\|_{p,\,q}$ is convex as a function of $(1/p, 1/q)$. In the complex case this is true for $p \geqslant 1, q \geqslant 1$. In the real case there is the additional restriction $(1/p) + (1/q') \geqslant 1$, which is the same as $q \geqslant p$. We summarize:

Theorem 4.52–A. *If* $A \in [l^p(n), l^q(m)]$, *then* $\log \|A\|_{p,\,q}$ *is a convex function of* $(1/p, 1/q)$ *for all* $p \geqslant 1$, $q \geqslant 1$ *in the complex case and for* $q \geqslant p \geqslant 1$ *in the real case.*

In order to display more directly the implication of this theorem, suppose $\psi(t)$ is a positive function such that $\log \psi(t)$ is convex when $0 \leqslant t \leqslant 1$. In particular, when $0 < t < 1$, the point $(t, \log \psi(t))$ on the graph of $\log \psi(t)$ is on or below the straight line joining the points $(0, \log \psi(0))$ and $(1, \log \psi(1))$. When the condition for this is worked out, it is found to be

$$\psi(t) \leqslant [\psi(0)]^{1-t}[\psi(1)]^t. \tag{4.52–3}$$

Suppose now that we are given pairs (p_1, q_1), (p_2, q_2) which are admissible under the conditions stated in Theorem 4.52–A. Let p and q be defined in terms of t by

$$\frac{1}{p} = \frac{1-t}{p_1} + \frac{t}{p_2}, \quad \frac{1}{q} = \frac{1-t}{q_1} + \frac{t}{q_2}, \qquad 0 < t < 1. \tag{4.52–4}$$

Let $\psi(t) = \|A\|_{p,\,q}$. Then (4.52–3) becomes

$$\|A\|_{p,\,q} \leqslant (\|A\|_{p_1,\,q_1})^{1-t}(\|A\|_{p_2,\,q_2})^t. \tag{4.52–5}$$

As t varies from 0 to 1, p varies from p_1 to p_2; likewise for q. Formula (4.52–5) expresses the content of Theorem 4.52–A. For example, if $p_1 = 1, q_1 = 2, p_2 = 2, q_2 = \infty$, we find that

$$p = \frac{2}{2-t}, \qquad q = \frac{2}{1-t} = \frac{2p}{2-p}.$$

In this case (4.52–5) becomes

$$\|A\|_{p,\,2p/(2-p)} \leqslant (\|A\|_{1,\,2})^{(2-p)/p}(\|A\|_{2,\,\infty})^{2(p-1)/p}.$$

Here p may vary from 1 to 2.

Now let us consider an infinite matrix (α_{ij}), $i, j = 1, 2, \cdots$. Let $A_{n,\,m}$ be the mapping of $l^p(n)$ into $l^q(m)$ defined by considering α_{ij} for $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant m$. It is easily proved that the infinite matrix defines a continuous linear mapping A of l^p into l^q if and only if $\|A_{n,m}\|_{p,q}$ is bounded as m and n vary and that, when such is the case,

$$\|A\|_{p,q} = \sup_{m,n} \|A_{n,m}\|_{p,q}. \tag{4.52–6}$$

We write $\|A\|_{p,q}$ rather than $\|A\|$, because we now wish to consider the possibility that A may be regarded as belonging to $[l^p, l^q]$ for various values of p and q.

Theorem 4.52–B. *Suppose that* (α_{ij}) *is an infinite matrix of scalars. Let* (p_1, q_1) *and* (p_2, q_2) *be admissible pairs according to the conditions of Theorem* 4.52–A. *Suppose the matrix defines an* A *belonging both to* $[l^{p_1}, l^{q_1}]$ *and to* $[l^{p_2}, l^{q_2}]$. *Then* A *belongs to* $[l^p, l^q]$ *for each* (p, q) *given by* (4.52–4), *and the norms in these various cases satisfy the inequality* (4.52–5).

The proof is immediate, by applying (4.52–5) to $A_{n,m}$ and taking note of the assertion made in connection with (4.52–6).

PROBLEMS

1. If the infinite matrix (α_{ij}) defines an operator $A \in [l^p]$, denote its norm by $\|A\|_p$. Show that, if the operator defined by the matrix belongs to $[l^1]$ and $[l^\infty]$, it belongs to l^p for $1 < p < \infty$, and $\|A\|_p \leqslant (\|A\|_1)^{1/p}(\|A\|_\infty)^{1-(1/p)}$. For a different approach to the case $p = 2$, see § 6.12.

2. Carry out the proof of (4.52–6).

4.6 Annihilators, Ranges, and Null Manifolds

We are now ready to take up the study of matters corresponding to the subject matter of §§ 1.9, 1.91, but with X', Y' replacing X^f, Y^f and A' replacing A^T. The differences are due mainly to the intervention of topological matters: the continuity of functionals and operators, and the distinction between closed and arbitrary linear manifolds.

The spaces X and Y referred to in this section are understood, without further mention, to be normed linear spaces. Completeness is not needed except in some of the problems. Many of the ideas and results apply to general topological linear spaces and especially to locally convex ones, but we shall not concern ourselves with the greater generality of such spaces.

In this section, and in applications based on it, the definitions of annihilators are as follows: If $G \subset X$, the annihilator of G in X' is the set

$$G^0 = \{x' \in X' : x'(x) = 0 \qquad \text{if } x \in G\}.$$

If $H \subset X'$, the annihilator of H in X is the set

$$^0H = \{x \in X : x'(x) = 0 \qquad \text{if } x' \in H\}.$$

It is assumed that G and H are nonempty.

The annihilator G^0 is a closed linear manifold in X'; it coincides with M^0, where M is the closed linear manifold generated by G. Similar remarks apply to ^0H.

In certain proofs in §§ 1.9, 1.91 we used Theorems 1.71–B and 1.71–C. For present purposes, in the corresponding circumstances, we can appeal to Theorems 4.3–D and 4.3–C. For later use we state the following theorems, leaving the proofs to the reader.

Theorem 4.6–A. *If $G \subset X$ and M is the closed linear manifold in X generated by G, then $^0(G^0) = M$. In particular, $^0(M^0) = M$ if M is any closed linear manifold in X.*

Theorem 4.6–B. *If $H \subset X'$ and N is the closed linear manifold in X' generated by H, then $N \subset (^0H)^0$. In particular, $N \subset (^0N)^0$ if N is any closed linear manifold in X'.*

A linear manifold N in X' will be called *saturated* if $N = (^0N)^0$. Such an N is necessarily closed. But a closed linear manifold is not necessarily saturated. To see this we refer to problem 1, § 4.32 and to the example at the end of § 1.9. If $X = (c)$, every $x' \in X'$ is representable in the form

$$x'(x) = \sum_0^\infty \alpha_{k+1}\xi_k, \qquad \text{with} \qquad \sum_1^\infty |\alpha_k| < \infty.$$

The space $(c)'$ is congruent in a natural way to l^1. Now let N be the set of those x' for which $\alpha_1 = 0$. This is a proper closed subspace of X'. But $^0N = (0)$ (argument given in § 1.9) and hence $(^0N)^0 = X'$. Thus N is not saturated.

In the case of any X there do exist saturated subspaces of X'; (0) and X' are examples. So is H^0, where H is any nonempty set in X.

In view of Theorem 4.6–B we note that, if N is any linear manifold in X', it is saturated if and only if $(^0N)^0 \subset N$. This is equivalent to

$$\text{complement of } N \subset \text{complement of } (^0N)^0.$$

Hence, a proper subspace N of X' is saturated if and only if to each $x' \in X' - N$ corresponds an $x \in {}^0N$ for which $x'(x) \neq 0$. Thus, a saturated subspace of X' is what has been called a *regularly closed* subspace (see Banach, **1**, pages 115–117). The notion of a saturated subspace is closely related to the weak topology $\mathscr{T}(X', X)$ for X' [$\mathscr{T}(X', X)$ is defined in the second paragraph after Theorem 3.81–B]. In fact, as we show later, in § 4.62, a subspace N of X' is saturated if and only if it is closed in the weak topology $\mathscr{T}(X', X)$.

We now turn to an examination of theorems analogous to those of § 1.91. When A is a continuous linear operator, the null manifold $\mathscr{N}(A)$ is a closed subspace; it is closed because it is the inverse image, by A, of a closed set. However, the range of A need not be closed.

Example. Let $X = Y = l^1$, and define $A \in [X]$ by $Ax = y$, where $y = \{\eta_k\}$, $x = \{\xi_k\}$, and $\eta_k = k^{-1}\xi_k$. Clearly $\mathscr{R}(A)$ contains all y's for which $\eta_k = 0$ except for a finite number of indices k, and so $\overline{\mathscr{R}(A)} = X$. But $\mathscr{R}(A) \neq X$; for instance, the vector y with $\eta_k = k^{-2}$ is in $X - \mathscr{R}(A)$, because $\xi_k = k^{-1}$ does not define an element of X. It follows that $\mathscr{R}(A)$ is not closed.

In the theorems which correspond to those of §1.91, it is the closure of the range of an operator which takes the place of the range itself. This is because of two things: (1) an annihilator is always closed, and (2) a linear manifold and its closure possess the same annihilator.

For later reference we state the following three theorems, leaving the proofs to the reader. It is assumed that $A \in [X, Y]$; X and Y need not be complete.

Theorem 4.6–C. $\{\overline{\mathscr{R}(A)}\}^0 = \mathscr{N}(A')$.

Theorem 4.6–D. $\overline{\mathscr{R}(A)} = {}^0\{\mathscr{N}(A')\}$.

Theorem 4.6–E. *$\overline{\mathscr{R}(A)} = Y$ if and only if $(A')^{-1}$ exists.*

At this point we are saying nothing about the continuity of $(A')^{-1}$. In § 4.7 we have some deeper theorems which relate the situation $\mathscr{R}(A) = Y$ to the existence and continuity of $(A')^{-1}$. These deeper theorems involve the issue of completeness.

When we exchange the roles of A and A' we do not get quite as much information as in the foregoing set of theorems. The following results are easily proved, and we leave proofs to the reader.

Theorem 4.6–F. ${}^0\{\overline{\mathscr{R}(A')}\} = \mathscr{N}(A)$.

Theorem 4.6–G. $\overline{\mathscr{R}(A')} \subset \{\mathscr{N}(A)\}^0$.

Theorem 4.6–H. *If $\overline{\mathscr{R}(A')} = X'$, then A^{-1} exists.*

Theorem 4.6–I. *If A^{-1} exists, then $({}^0\{\overline{\mathscr{R}(A')}\})^0 = X'$.*

In connection with Theorem 4.6–G we observe that, if $\overline{\mathscr{R}(A')}$ is saturated in X', we have the equality $\overline{\mathscr{R}(A')} = \{\mathscr{N}(A)\}^0$. This happens, for instance, if X is norm reflexive or if $\overline{\mathscr{R}(A')}$ is finite dimensional (see problems 1, 2 at the end of this section). Likewise, when A^{-1} exists and we have reason to know that $\overline{\mathscr{R}(A')}$ is saturated, we can conclude from Theorem 4.6–I that $\overline{\mathscr{R}(A')} = X'$.

PROBLEMS

1. If X is norm reflexive, every closed subspace of X' is saturated.

2. A finite-dimensional subspace of X' is saturated (see Theorem 3.5–C).

3. If $\{\alpha_n\}$ is a sequence of scalars with $\sup_n |\alpha_n| < \infty$, let $X = l^1$, and let A be the operator $A \in [l^1]$ defined by $Ax = y$, where $\eta_k = \alpha_k \xi_k$. Then $\|A\| = \sup_n |\alpha_n|$. The inverse A^{-1} exists if and only if $\alpha_n \neq 0$ for every n, and then $\overline{\mathscr{R}(A)} = X$. Show that $\mathscr{R}(A) = X$ and A^{-1} is continuous if and only if $\inf_n |\alpha_n| > 0$.

4. If X is complete and X' is norm reflexive, so is X. Let J_0 and J_1 be the canonical mappings of X and X' into X'' and X''', respectively. Show that $(J_1)^{-1} = J_0'$, as a result of the reflexivity of X'. Then use Theorem 4.6–E.

5. Let M be a closed subspace of the normed linear space X. Let $F = X/M$ (normed as in § 3.14). Then F' is congruent to M^0. The congruence is established by the mapping $Tf = x'$, where $f \in F'$ and $x'(x) = f([x])$, $x \in X$, $x' \in X'$, $[x] \in X/M$. It is to be shown that T is linear and isometric, with range M^0.

6. Suppose M is a closed subspace of the norm-reflexive space X. Then the quotient space X/M is norm reflexive. The argument is that X' is norm reflexive and hence also, in turn, M^0, $(X/M)'$, and X/M. See the two preceding problems and problem 4, § 4.5.

7. Suppose that X is complete and that $\mathscr{R}(A)$, as a normed linear space on its own account, is complete [as it is, for example, if Y is complete and $\mathscr{R}(A)$ is closed]. Then $\mathscr{R}(A') = \{\mathscr{N}(A)\}^0$. In view of Theorem 4.6–G it suffices to show that $\{\mathscr{N}(A)\}^0 \subset \mathscr{R}(A')$. If $x' \in \{\mathscr{N}(A)\}^0$ and $y \in \mathscr{R}(A)$, then $x'(x)$ has the same value for each x such that $y = Ax$. Hence define $f(y) = x'(x)$, using such an x. Use Theorem 4.2–G(5) to show that $f \in \{\mathscr{R}(A)\}'$. Then extend f to obtain $y' \in Y'$ and show that $A'y' = x'$.

4.61 Weak Compactness in Normed Linear Spaces

The use of weak topologies sheds light on various questions which arise in relation to normed linear spaces. In this section we consider compactness, using weak topologies.

Throughout the section X denotes a normed linear space, X' its normed conjugate, and J the canonical mapping of X into X''. We shall be concerned with the topology $\mathscr{T}(X, X')$ for X and the topology $\mathscr{T}(X', X)$ for X' (see the third paragraph of § 4.41). We also consider the topology $\mathscr{T}(X'', X')$ for X''.

The first main result is the following:

Theorem 4.61–A. *Suppose $S \subset X'$. Let S be bounded with respect to the norm in X' and closed with respect to $\mathscr{T}(X', X)$. Then S is compact with respect to $\mathscr{T}(X', X)$. In particular, $\{x' : \|x'\| \leqslant 1\}$ is compact in this sense.*

PROOF. Suppose $C = \sup_{x \in S} \|x'\|$. For each $x \in X$ let $A_x = \{\alpha : |\alpha| \leqslant C\|x\|\}$. Let $A = \prod_{x \in X} A_x$ (the Cartesian product as in § 2.5). Then A is compact, by Theorem 2.5–A. Now, any element $x' \in S$ is determined by the set of values $x'(x)$, $x \in X$; since $|x'(x)| \leqslant \|x'\|\,\|x\| \leqslant C\|x\|$, we see that $x' \in A$ or $S \subset A$. Moreover, it is easily verified that the topology induced on S by $\mathscr{T}(X', X)$ is the same as the topology induced on S by the Cartesian product topology for A. Hence, for the first assertion in the theorem, it suffices to prove that S is closed as a subset of A (see Theorem 2.2–B). Suppose f is an element of the closure of S in A. Consider any $\epsilon > 0$ and any $x, y \in X$. The set of all $g \in A$ such that

$$|g(x) - f(x)| < \epsilon, \qquad |g(y) - f(y)| < \epsilon,$$

and

$$|g(x + y) - f(x + y)| < \epsilon,$$

is a neighborhood of f in A. This neighborhood contains some point $x' \in S$, and, since x' is linear, we have

$$|f(x + y) - f(x) - f(y)| \leqslant |f(x + y) - x'(x + y)| + |x'(x) - f(x)| + |x'(y) - f(y)| < 3\epsilon,$$

whence it follows that $f(x + y) = f(x) + f(y)$. In a similar way we prove that $f(\beta x) = \beta f(x)$. Now, by the fact that $f \in A$, we know that $f(x) \in A_x$ or $|f(x)| \leqslant C\|x\|$. Hence $f \in X'$. Then $f \in S$, for S is closed in X' relative to $\mathscr{T}(X', X)$. This proves that S is closed in A and hence that S is compact relative to $\mathscr{T}(X', X)$. The last assertion of the theorem follows from the fact that $\{x' : \|x'\| \leqslant 1\}$ is closed in X' relative to $\mathscr{T}(X', X)$ (see problems 9 and 10 in § 4.4).

The next main result is Theorem 4.61–C, for the proof of which we need the following lemma:

Lemma 4.61–B. *Consider $S \subset X$* and *$S'' \subset X''$ defined as follows:*

$$S = \{x : \|x\| \leqslant 1\}, \quad S'' = \{x'' : \|x''\| \leqslant 1\}.$$

Also consider the canonical image $J(S)$ *of* S *in* X''. *This set* $J(S)$ *is dense in* S'' *with respect to the topology* $\mathscr{T}(X'', X')$.

PROOF. Suppose $x_0'' \in S''$, and suppose $\epsilon > 0$, $x_1', \cdots, x_n' \in X'$. We shall produce an $x_0 \in S$ such that Jx_0 is in the neighborhood of x_0'' defined by ϵ, $x_1', \cdots, x_n'$. That is, x_0 will be such that

$$|x_i'(x_0) - x_0''(x_i')| < \epsilon, \qquad i = 1, \cdots, n. \tag{4.61–1}$$

Let M be the set of all $x \in X$ such that $x_i'(x) = 0$ if $i = 1, \cdots, n$. Then M is a subspace of X, closed in the norm topology. By Theorem 3.5–C, M^0 is the subspace of X' generated by $x_1', \cdots, x_n'$ (it suffices to work with a maximal linearly independent subset of $x_1', \cdots, x_n'$). Consider the quotient space X/M. The fact that the dimension of M^0 does not exceed n implies the same property for X/M. In fact, $(X/M)'$ is congruent to M^0 (problem 5, § 4.6), so that $(X/M)'$, and so also X/M, have the same finite dimension as M^0. Now, the space X/M, being finite dimensional, is reflexive. We utilize the congruence of $(X/M)'$ and M^0 under the mapping $x' = T(V')$, where $V' \in (X/M)'$, $x' \in M^0$, and $x'(x) = V'([x])$. Using the given $x_0'' \in S''$, we define $V'' \in (X/M)''$ by $V''(V') = x_0''(TV')$. Evidently $\|V''\| \leqslant 1$, because T is norm-preserving. Because of the reflexivity of X/M there exists $V \in X/M$ such that $V''(V') = V'(V)$ for every V', and $\|V\| = \|V''\| \leqslant 1$. Now choose $C > \sup_i \|x_i'\|$. By the definition of norms in X/M, there exists $x \in V$ such that $\|x\| < \|V\| + \epsilon/C \leqslant 1 + (\epsilon/C)$. By the definition of T, $x'(x) = V'(V)$, where $x' = TV'$. Therefore $x'(x) = x_0''(x')$ for every $x' \in M^0$. Now let $x_0 = Cx/(C + \epsilon)$. We see that $\|x_0\| < 1$ and $\|x - x_0\| = \epsilon\|x\|/(C + \epsilon) < \epsilon/C$. Also, if $x' \in M$,

$$|x_0''(x') - x'(x_0)| = |x'(x) - x'(x_0)| \leqslant \|x'\| \, \|x - x_0\| < \epsilon\|x'\|/C;$$

therefore, by the definition of C, it follows that (4.61–1) holds. This proves the theorem.

We observe that, if X is given the topology $\mathscr{T}(X, X')$ and the canonical image $J(X)$ of X in X'' is given the topology induced on it by $\mathscr{T}(X'', X')$, then J is a linear homeomorphism of X onto $J(X)$. This is clear directly from the way in which the bases at 0 are defined for the topologies in question.

Theorem 4.61–C. *The normed linear space* X *is norm reflexive if and only if the set* $S = \{x : \|x\| \leqslant 1\}$ *is compact with respect to the weak topology* $\mathscr{T}(X, X')$.

PROOF. If S is compact, $J(S)$ is compact relative to the topology $\mathscr{T}(X'', X')$, since J is continuous, as noted above. Hence, by Theorem

2.3–B, $J(S)$ is closed in X''. It then follows from Lemma 4.61–B that $J(S) = S''$, for S'' is closed with respect to $\mathscr{T}(X'', X')$ (see problem 10, § 4.4). It now follows easily that $J(X) = X''$, so that X is norm reflexive.

On the other hand, if X is norm reflexive, $J^{-1}(S'') = S$. Now S'' is compact relative to $\mathscr{T}(X'', X')$ (Theorem 4.61–A), and hence S is compact relative to $\mathscr{T}(X, X')$, since J^{-1} is continuous.

In the older literature dealing with Banach spaces, before the systematic development of the weak topologies, a good deal of attention was paid to the concept of weak convergence of sequences, even though the weak topologies themselves were not investigated. In particular, the following concept was used (though with different terminology):

Definition. In the normed linear space X, a set S is called conditionally sequentially X'-weakly compact if every sequence from S contains a subsequence which is X'-weakly convergent to some limit in X.

We see from Theorem 4.41–B that if X is norm reflexive, every bounded set in X is conditionally sequentially X'-weakly compact. This proposition has a valid converse, as follows:

Theorem 4.61–D. *If X is a Banach space such that every bounded sequence contains an X'-weakly convergent subsequence, then X is norm reflexive.*

We shall not give a complete proof of this theorem. It was first proved under the additional assumption that X is separable (Banach **1**, pages 189–191). After a great deal of development of the subject of weak topologies it was proved that, *if X is a Banach space and S is a set in X which is closed relative to the topology $\mathscr{T}(X, X')$ and conditionally sequentially X'-weakly compact, then S is compact relative to $\mathscr{T}(X, X')$* (Eberlein, **1**). Theorem 4.61–D is a consequence of this result, in view of Theorem 4.61–C and the fact that $\{x : \|x\| \leqslant 1\}$ is closed relative to $\mathscr{T}(X, X')$ (see problem 9, § 4.4).

A partial outline of the proof of Eberlein's result runs as follows: consider the canonical image $J(S)$ of S in X''. Since S is bounded (in the norm sense), so is $J(S)$. The difficult thing to prove is that $J(S)$ is closed relative to the topology $\mathscr{T}(X'', X')$. Once this is done, Theorem 4.61–A shows that $J(S)$ is compact relative to $\mathscr{T}(X'', X')$. Finally, S itself is compact relative to $\mathscr{T}(X, X')$, because J^{-1} is continuous in the relevant topologies. The proof that $J(S)$ is closed relative to $\mathscr{T}(X'', X')$ makes use of the important theorem 4.62–B, which we state, but do not prove, in the next section. Use is also made of the interesting fact (see problem 4, § 3.81) that, if K is a convex set in X, its closure in the norm topology coincides with its closure in the topology $\mathscr{T}(X, X')$.

For further information about compactness relative to the topology $\mathscr{T}(X, X')$, see Dunford and Schwartz, **1**, Chapter V, § 6.

PROBLEMS

1. Let X be a normed linear space, and suppose $x_1', \cdots, x_n' \in X'$. Suppose $\alpha_1, \cdots, \alpha_n$ are scalars, that $B > 0$, and that $|\lambda_1\alpha_1 + \cdots + \lambda_n\alpha_n| \leqslant B\|\lambda_1 x_1' + \cdots + \lambda_n x_n'\|$ for every set of scalars $\lambda_1, \cdots, \lambda_n$. Then, to each $\delta > 0$ corresponds some x with $\|x\| \leqslant B + \delta$ and $x_i'(x) = \alpha_i$, $i = 1, \cdots, n$. Outline of proof: Let T be the mapping of X into $l^2(n)$ defined by $Tx = [x_1'(x), \cdots, x_n'(x)]$. Let $K = \{x : \|x\| \leqslant B + \delta\}$. Show that there exists a linear functional on $l^2(n)$ with value 1 at $(\alpha_1, \cdots, \alpha_n)$ and of absolute value not exceeding 1 on $T(K)$. If the value of this functional at $(\xi_1, \cdots \xi_n)$ is $\lambda_1\xi_1 + \cdots + \lambda_n\xi_n$, deduce the contradiction $\lambda_1\alpha_1 + \cdots + \lambda_n\alpha_n = 0$. This is Y. Mimura's proof of a result due originally to E. Helly. See Kakutani, **1**, page 171.

2. Use the result in problem 1 to arrive at (4.61–1) in a different way, thus giving an alternative proof of Lemma 4.61–B. Take $B = \|x_0''\|$, $\alpha_i = x_0''(x_i')$. Then x_0 can be chosen as a suitable multiple of the x obtained by using problem 1.

3. A normed linear space X is called *uniformly* convex if to each $\epsilon > 0$ corresponds $\delta(\epsilon) > 0$ such that $\|x + y\| \leqslant 2(1 - \delta(\epsilon))$ when $\|x\| \leqslant 1$, $\|y\| \leqslant 1$ and $\|x - y\| \geqslant \epsilon$. This concept is due to Clarkson, **1**. Every uniformly convex Banach space is norm reflexive. This has been proved in different ways by Milman, **1**, Pettis, **1**, and Kakutani, **1**. Kakutani's argument depends on problem 1. It will suffice to show that, if $x_0'' \in X''$ and $\|x_0''\| = 1$, then $x_0'' \in \mathscr{R}(J)$, J being the canonical mapping of X into X''. Choose $x_k' \in X'$ so that $\|x_k'\| = 1$ and $x_0''(x_k') > 1 - (1/k)$. Applying problem 1 with $B = 1$ and $\alpha_i = x_0''(x_i')$, $i = 1, \cdots, n$, we obtain $x_n \in X$ so that $\|x_n\| < 1 + (1/n)$ and $x_i'(x_n) = x_0''(x_i')$, $i = 1, \cdots, n$. Use uniform convexity to deduce a contradiction if one assumes that $\{x_n\}$ is not a Cauchy sequence. Then $x_n \rightarrow x_0$, and it follows easily that $\|x_0\| = 1$, $x_n'(x_0) = x_0''(x_n)$ for $n = 1, 2, \cdots$. Using uniform convexity again, it can be seen that these last mentioned conditions determine x_0 uniquely. Now suppose $x_0' \in X'$. Proceeding from problem 1 again, we obtain $y_n \in X$ so that $\|y_n\| < 1 + (1/n)$ and $x_i'(y_n) = x_0''(x_i')$ if $i = 0, 1, \cdots, n$. Then $\{y_n\}$ is a Cauchy sequence with limit y_0 such that $\|y_0\| = 1$, $x_n'(y_0) = x_0''(x_n')$ if $n = 0, 1, \cdots$. By the uniqueness, $y_0 = x_0$ and hence $Jx_0 = x_0''$.

4.62 Saturated Subspaces of a Conjugate Space

If X is a normed linear space, the topology $\mathscr{T}(X', X)$ is weaker than $\mathscr{T}(X', X'')$, which in turn is weaker than the topology generated by the norm in X'. A subspace of X' is closed relative to $\mathscr{T}(X', X'')$ if and only if it is closed in the norm topology (by Theorem 3.81–C). Because $\mathscr{T}(X', X)$ is weaker than $\mathscr{T}(X', X'')$, it follows that a subspace of X' which is closed relative to $\mathscr{T}(X', X)$ is also closed relative to $\mathscr{T}(X', X'')$.

The converse is not true in general; in fact, as we shall see, if X has the property, that every subspace of X' which is closed relative $\mathscr{T}(X', X'')$ is also closed relative to $\mathscr{T}(X', X)$, then X must be norm reflexive.

Theorem 4.62–A. *A subspace M of X' is closed relative to $\mathscr{T}(X', X)$ if and only if it is saturated.*

PROOF. Suppose M is saturated and that $x_0' \in X' - M$. Then there exists some $x_1 \in {}^0M$ such that $x_0'(x_1) \neq 0$. Choose ϵ so that $0 < \epsilon < |x_0'(x_1)|$. Then $\{x' : |x'(x_1) - x_0'(x_1)| < \epsilon\}$ is a neighborhood of x_0' in the topology $\mathscr{T}(X', X)$. This neighborhood does not intersect M, and so M is closed.

Now suppose that M is closed relative to $\mathscr{T}(X', X)$. We can assume $M \neq X'$, for X' is saturated. Suppose $x_0' \in X' - M$. The fact that M is closed means that there exists $\epsilon > 0$ and $x_1, \cdots, x_n \in X$ such that, if $x' \in M$, then

$$|x'(x_i) - x_0'(x_i)| \geqslant \epsilon, \qquad i = 1, 2, \cdots, n. \tag{4.62–1}$$

Let T be the linear mapping of X' into $l^2(n)$ defined by $Tx' = \{x'(x_1), \cdots, x'(x_n)\}$. Then $T(M)$ is a closed subspace of $l^2(n)$, and it does not contain Tx_0'. Hence there exists a linear functional on $l^2(n)$ which has the value 0 on $T(M)$ and a nonzero value at Tx_0'. Let $\sum_{i=1}^{n} \alpha_i \xi_i$ be the value of this functional at the point $(\xi_1, \cdots, \xi_n)$. Define $x_0 = \sum_{i=1}^{n} \alpha_i x_i$. Then $x'(x_0) = 0$ if $x' \in M$, and $x_0'(x_0) \neq 0$. This proves that M is saturated. The proof is now complete.

We have not the space or the need in this book to push very far with a study of saturated subspaces. However, we mention without proof the following very important theorem.

Theorem 4.62–B. *Let X be a Banach space. Let M be a subspace of X' such that $M \cap \{x' : \|x'\| \leqslant 1\}$ is compact relative to $\mathscr{T}(X', X)$. Then M is saturated.*

In view of Theorem 4.61–A, the word "compact" may be replaced by "closed" in the statement of Theorem 4.62–B. This theorem, due to Dieudonné (**1**, Theorème 23), is closely related to Lemma 2, page 119 in the book **1** of Banach.

PROBLEMS

1. Suppose $x'' \in X''$, and let $M = \{x' : x''(x') = 0\}$. Then $x'' \in J(X)$ if and only if M is saturated. In proving the "if" part use Theorem 3.5–C.

2. The space X is norm reflexive if every subspace of X' which is closed in the norm topology is saturated. See problem 1 of § 4.6 for the converse.

3. Use Theorem 4.62–B to prove the "if" part of Theorem 4.61–C without using Lemma 4.61–B. Using the notation of these propositions, observe that $J(S) = J(X) \cap S''$, and deduce that $J(X)$ is saturated, whence $J(X) = X''$.

4. Let M be a subspace of X', and let $\bar{M}$ be its closure in the topology $\mathscr{T}(X', X)$. Then $\bar{M} = X'$ if and only if M is total. Note that "M is total" can be expressed in the form ${}^0M = (0)$.

4.7 Theorems about Continuous Inverses

Throughout this section we assume that X and Y are normed linear spaces and that $A \in [X, Y]$, whence $A' \in [Y', X']$. In some cases we assume completeness for X or Y; these assumptions are made explicit in each case. The conjugate spaces X', Y' are complete, in any event.

Theorem 4.7–A. *$\mathscr{R}(A') = X'$ if and only of A^{-1} exists and is continuous.*

PROOF. Suppose that $\mathscr{R}(A') = X'$ but that A does not have a continuous inverse. Then, by Theorem 3.1–B, there is a sequence $\{x_n\}$ such that $x_n \neq 0$ and $\|Ax_n\|/\|x_n\| \to 0$. Let

$$\alpha_n = \max\left\{\left(\frac{\|Ax_n\|}{\|x_n\|}\right)^{\frac{1}{2}},\ n^{-\frac{1}{2}}\right\}, \qquad u_n = \frac{x_n}{\alpha_n\|x_n\|}.$$

Then $\|u_n\| = \dfrac{1}{\alpha_n} \to \infty$, and

$$\|Au_n\| = \frac{\|Ax_n\|}{\alpha_n\|x_n\|} \leqslant \left(\frac{\|Ax_n\|}{\|x_n\|}\right)^{\frac{1}{2}} \to 0.$$

But then $y'(Au_n) \to 0$ for each $y' \in Y'$. Since $\mathscr{R}(A') = X'$ and $y'(Au_n) = \langle Au_n, y'\rangle = \langle u_n, A'y'\rangle$, this implies that $x'(u_n) \to 0$ for each $x' \in X'$. But then, Theorem 4.4–D*a* implies that $\{\|u_n\|\}$ is bounded, which is a contradiction.

Now suppose, conversely, that A has a continuous inverse. If x' is fixed in X' and y is variable in $\mathscr{R}(A)$, $x'(A^{-1}y)$ defines a continuous linear functional on $\mathscr{R}(A)$. This functional can be extended to all of Y, the extension being a member of Y' (Theorem 4.3–A). That is, there exists $y' \in Y'$ such that $y'(y) = x'(A^{-1}y)$ when $y \in \mathscr{R}(A)$. Thus $y'(Ax) = x'(x)$ for each $x \in X$. This means that $x' = A'y'$. Therefore $\mathscr{R}(A')$ is all of X'. The proof is now complete.

Theorem 4.7–B. *Suppose that Y is complete and that $\mathscr{R}(A) = Y$. Then A' has a continuous inverse.*

PROOF. The argument is very much like that in the first part of the proof of Theorem 4.7–A. If A' did not have a continuous inverse, we should be able to assert the existence of a sequence $\{y_n'\}$ such that $\|y_n'\| \to \infty$ and $\|A'y_n'\| \to 0$. Then $\langle x, A'y_n'\rangle \to 0$ for each $x \in X$. Since $\langle x, A'y_n'\rangle = \langle Ax, y_n'\rangle$ and $\mathscr{R}(A) = Y$, this means that $y_n'(y) \to 0$ for each $y \in Y$. Since Y is complete, we see by Theorem 4.4–D*b* that $\{\|y_n'\|\}$ is bounded. This is a contradiction, so the theorem is proved.

The next theorem is in the nature of a converse of Theorem 4.7–B. But the proof is more complicated, and we need to assume that X is complete.

Theorem 4.7–C. *Suppose that X is complete and that A' has a continuous inverse. Then $\mathscr{R}(A) = Y$. Moreover, it follows that Y is complete. Finally, A^{-1} is continuous, if it exists.*

PROOF. For each $\alpha > 0$ let $B_\alpha = \{x : \|x\| \leqslant \alpha\}$. We first prove that the set $\overline{A(B_\alpha)}$ contains a neighborhood of 0 in Y. This part of the proof does not require X to be complete. The set B_α is convex and balanced, so the same is true of $A(B_\alpha)$ and $\overline{A(B_\alpha)}$. Suppose that, for some α, $\overline{A(B_\alpha)}$ does not contain a neighborhood of 0. Then there exists a sequence $\{y_n\}$ in Y such that $y_n \to 0$ and $y_n \notin \overline{A(B_\alpha)}$. Let d_n be the distance between y_n and $\overline{A(B_\alpha)}$. If $y_0 \in \overline{A(B_\alpha)}$, consider the closed sphere $\{y : \|y - y_0\| \leqslant d_n/2\}$, and (for a fixed n) the union S_n of all such spheres corresponding to all choices of y_0. Let $V_n = \bar{S}_n$. By a very simple argument the convexity of $\overline{A(B_\alpha)}$ implies that of S_n, and hence V_n is convex. Likewise, V_n is balanced. Since $0 \in A(B_\alpha)$, V_n contains 0 as an interior point. Hence $\epsilon y_n \in V_n$ if ϵ is sufficiently small. However, $y_n \notin V_n$; therefore, if ϵ_n is the largest positive value of ϵ such that $\epsilon y_n \in V_n$, we know that $\epsilon_n < 1$ and that $\epsilon_n y_n$ is on the boundary of V_n. By Theorem 3.7–D there exists an element $y' \in Y'$ such that $y'(\epsilon_n y_n) = 1$ and $|y'(y)| \leqslant 1$ if $y \in V_n$. In particular, $|y'(Ax)| \leqslant 1$ if $\|x\| \leqslant \alpha$, and $|y'(Ax)| \leqslant \alpha^{-1}$ if $\|x\| \leqslant 1$. This means that $\|A'y'\| \leqslant \alpha^{-1}$. Now, the fact that A' has a continuous inverse means that $\|(A')^{-1}x'\| \leqslant m\|x'\|$ for each $x' \in \mathscr{R}(A')$, where m is some positive constant. If $x' = A'y'$ we then have $\|y'\| \leqslant m\alpha^{-1}$. Then $1 = y'(\epsilon_n y_n) \leqslant \|y'\|\,\|\epsilon_n y_n\| < m\alpha^{-1}\|y_n\|$, because $0 < \epsilon_n < 1$. But $\|y_n\| \geqslant \alpha/m$ contradicts the fact that $y_n \to 0$. This shows that $\overline{A(B_\alpha)}$ contains a neighborhood of 0.

The rest of the proof is much like the last part of the argument used in proving assertion 1 of Theorem 4.2–G. There is some $\beta > 0$ such that $\{y : \|y\| \leqslant \beta\} \subset \overline{A(B_\alpha)}$. Choose the values $1, 2^{-1}, 2^{-2}, \cdots, 2^{-n}, \cdots$ for

α, and denote corresponding values of β by $\beta_0, \beta_1, \cdots, \beta_n, \cdots$. Choose any y_0 for which $\|y_0\| \leqslant \beta_0$. Then $y_0 \in \overline{A(B_\alpha)}$ for $\alpha = 1$, and hence there exists an element $y_1 = Ax_1$, with $\|x_1\| \leqslant 1$, such that $\|y_0 - y_1\| \leqslant \beta_1$. This implies $y_0 - y_1 \in \overline{A(B_\alpha)}$ with $\alpha = 2^{-1}$. By induction we obtain elements $y_n = Ax_n$, with $\|x_n\| \leqslant 1/2^{n-1}$, such that $\|y_0 - y_1 - \cdots - y_n\| \leqslant \beta_n$, and hence $y_0 - y_1 - \cdots - y_n \in \overline{A(B_\alpha)}$ with $\alpha = 2^{-n}$. Now $\|Ax\| \leqslant \|A\|\alpha$ if $\|x\| \leqslant \alpha$. Therefore $\|y\| \leqslant \|A\|\alpha$ if $y \in \overline{A(B_\alpha)}$. This shows that $\beta_n \to 0$, and hence that

$$y_0 = \sum_1^\infty y_n.$$

Since X is complete and $\sum_1^\infty \|x_n\|$ is convergent, the series $\sum_1^\infty x_n$ converges and defines an element $x_0 \in X$. Moreover, $Ax_0 = \sum_1^\infty Ax_n = y_0$. This shows that $\{y : \|y\| \leqslant \beta_0\} \subset \mathscr{R}(A)$; it follows by homogeneity that $\mathscr{R}(A) = Y$.

It remains to prove that Y is complete. Let $\hat{Y}$ be the completion of Y, and define $\hat{A} \in [X, \hat{Y}]$ by writing $\hat{A}x = Ax$ if $x \in X$. As we remarked in § 4.3, we can identify $\hat{Y}'$ with Y'; hence we can identify $(\hat{A})'$ with A'. It follows that $(\hat{A})'$ has a continuous inverse, and so, by what has already been proved, we know that $\mathscr{R}(\hat{A}) = \hat{Y}$. But $\mathscr{R}(\hat{A}) = \mathscr{R}(A)$, and so $\hat{Y} = Y$; it follows that Y is complete. The fact that A^{-1} is continuous, if it exists, now follows from Theorem 4.2–H.

4.71 The States of an Operator and its Conjugate

The theorems in § 4.7, together with some of those in §§ 4.2, 4.6, make it possible for us to tabulate in diagrammatic form a large number of interesting and useful implications involving a bounded linear operator and its conjugate. In order to tabulate information concisely, we make a ninefold classification of what we shall call the *state* of an operator. If $A \in [X, Y]$, we list three possibilities for $\mathscr{R}(A)$, labelling them I, II, III, as follows:

I. $\mathscr{R}(A) = Y$.
II. $\overline{\mathscr{R}(A)} = Y$, but $\mathscr{R}(A) \neq Y$.
III. $\overline{\mathscr{R}(A)} \neq Y$.

As regards A^{-1} we also list three possibilities, labelling them 1, 2, 3:

1. A^{-1} exists and is continuous.
2. A^{-1} exists but is not continuous.
3. A^{-1} does not exist.

By combining these possibilities in all possible ways we obtain nine different situations. For instance, it may be that $\overline{\mathscr{R}(A)} = Y$, $\mathscr{R}(A) \neq Y$, and that A^{-1} exists but is discontinuous. We shall describe this as state II_2 for A; alternatively, we shall say that A is in state II_2. There are nine states in all: I_1, I_2, I_3, II_1, $\cdots$, III_2, III_3.

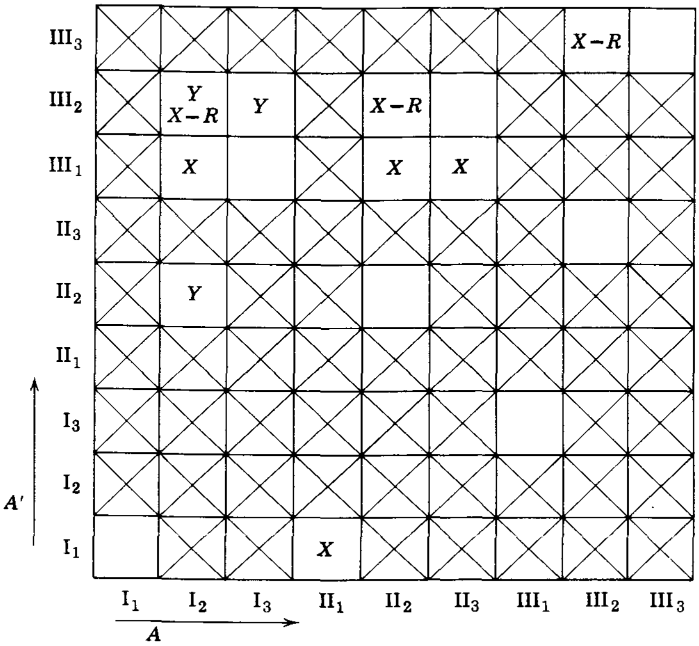
The notion of the state of an operator can be applied, in particular, to A'. Here A' replaces A and X' replaces Y in the description of the states; to say that A' is in state I_3 means that $\mathscr{R}(A') = X'$ and A' does not have an inverse.

It is possible for A to be in any one of the nine states if no restrictions are placed on X and Y. But, if we assume that X and Y are both complete, it is impossible for A to be in state I_2; this is a consequence of Theorem 4.2–H. Since X' and Y' are always complete, this shows that A' can never be in state I_2. Likewise, if X is complete, A can never be in state II_1. For, it follows by Theorem 4.2–B that A is closed; hence, if X is complete and A has a continuous inverse, Theorem 4.2–E shows that $\mathscr{R}(A)$ is closed, so that $\overline{\mathscr{R}(A)} = Y \neq \mathscr{R}(A)$ is impossible. Applying this result to A' (noting that Y' is complete), we see that A' can never be in state II_1.

Next we define the *state of the pair* (A, A'); this is the ordered pair of the states of A and A', respectively, with the state of A listed first. We shall now proceed to construct a diagram which shows the states of the pair (A, A') which are demonstrated to be impossible because of various theorems in §§ 4.2, 4.6, and 4.7. This diagram is a large square, divided into eighty-one smaller squares arranged in rows and columns. We label each column by a state symbol placed at the bottom; this denotes a certain state for A. We label each row by a state symbol placed at the left; this denotes a certain state for A'. The square which is the intersection of a certain column and row denotes a state of the pair (A, A') [the intersection of the I_2 column and III_1 row denoting the state $(\mathrm{I}_2, \mathrm{III}_1)$, and so on].

We cross out a small square by its diagonals if the corresponding state is impossible by virtue of our theorems, *without requiring X or Y to be complete*. The reader should now examine the state diagram, as it appears on page 237. We shall explain briefly the reasons for the crossing out of those squares which are crossed out. All squares in the I_2 and II_1 rows are crossed out, since A' can never be in these states; this was

THE STATE DIAGRAM



A'	I_1	I_2	I_3	II_1	II_2	II_3	III_1	III_2	III_3
III_3	☒	☒	☒	☒	☒	☒	☒	$X-R$	
III_2	☒	Y $X-R$	Y	☒	$X-R$		☒	☒	☒
III_1	☒	X		☒	X	X	☒	☒	☒
II_3	☒	☒	☒	☒	☒	☒	☒		☒
II_2	☒	Y	☒	☒		☒	☒	☒	☒
II_1	☒	☒	☒	☒	☒	☒	☒	☒	☒
I_3	☒	☒	☒	☒	☒	☒		☒	☒
I_2	☒	☒	☒	☒	☒	☒	☒	☒	☒
I_1		☒	☒	X	☒	☒	☒	☒	☒

A

X	Impossible with X complete
Y	Impossible with Y complete
$X-R$	Impossible with X norm reflexive

explained earlier. Theorem 4.6–E says that, if A is in a I or II state, A' must be in a 1 or 2 state, and vice versa. Thus any state of the pair with I or II for A and 3 for A' is impossible; also, III for A and 1 or 2 for A' is impossible. Theorem 4.6–H says that a state of the pair with 3 for A and I or II for A' is impossible. Theorem 4.7–A says that we have 1 for A if and only if we have I for A'. This eliminates states with 1 for A and II or III for A' and also those with 2 or 3 for A and I for A'. All but sixteen of the eighty-one squares are crossed out as a result of the foregoing considerations. These sixteen remaining states for the pair (A, A') are all possible, if no restrictions are placed on X or Y; this has been shown by the construction of examples (see Taylor and Halberg, **1**). Some of these examples are considered in the problems. Many implications can now be read from the state diagram. Samples: The state of A is III_1 if and only if the state of A' is I_3; if the state of A is III_2, the state of A' is either II_3 or III_3.

Next we consider the effect of assuming that X is complete. Theorem 4.7–C tells us, in this case, that any state except I_1 and I_3 for A is impossible if we have 1 for A'. Of the sixteen states previously not crossed out, this rules out $(\mathrm{I}_2, \mathrm{III}_1)$, $(\mathrm{II}_1, \mathrm{I}_1)$, $(\mathrm{II}_2, \mathrm{III}_1)$, and $(\mathrm{II}_3, \mathrm{III}_1)$. These exclusions are indicated on the state diagram by putting the letter X in the corresponding square.

Now consider the effect of assuming that Y is complete. Theorem 4.7–B says, in this case, that I for A makes 2 and 3 impossible for A'. This rules out $(\mathrm{I}_2, \mathrm{II}_2)$, $(\mathrm{I}_2, \mathrm{III}_2)$, and $(\mathrm{I}_3, \mathrm{III}_2)$, in addition to some states already crossed out. These exclusions are indicated by putting a Y in the appropriate square on the state diagram.

With X and Y both assumed to be complete, there are nine states still left as apparently possible for the pair (A, A'). Examples exist which show that these states are actually possible.

If we assume that X is norm reflexive, but make no restrictions on Y, Theorem 4.6–I and problem 1 of § 4.6 tell us that we must have I or II for A' if we have 1 or 2 for A. Hence, if X is norm reflexive, in addition to the states ruled out because X is complete, this rules out $(\mathrm{I}_2, \mathrm{III}_2)$, $(\mathrm{II}_2, \mathrm{III}_2)$ and $(\mathrm{III}_2, \mathrm{III}_3)$ from the sixteen states which are not crossed out by their diagonals. These three exclusions are indicated by $X - R$ in the corresponding square on the state diagram.

It is now interesting to observe, on the state diagram, that when X is norm reflexive and Y is complete, there are seven state squares left blank. Also, there is at most one blank square in any one row or column. Hence, with these conditions on X and Y, the state of either A or A' determines uniquely the state of the other. These seven states for the pair (A, A') do actually occur; examples can be constructed with $X = Y = l^2$.

PROBLEMS

In the first nine problems are given examples of operators A which provide illustrations of various states for the pair (A, A'). In these examples X and Y are chosen from the spaces $l^p(1 \leqslant p < \infty)$ and (c_0). In each of these spaces we denote the vector $(0, \cdots 0, 1, 0, \cdots)$ (with 1 in the k^{th} place) by u_k. In l^p, with $1 \leqslant p < \infty$, and (c_0) the set $\{u_k\}$ is a countable basis, and in defining A on these spaces it suffices to define Au_k for each k. If X and Y are both spaces in which $\{u_k\}$ is a countable basis, A' is determined by the transpose of the infinite matrix which represents A. In each problem it is left for the reader to show that the state of the pair (A, A') is as indicated.

1. $X = Y = l^2$. $Au_k = 2^{1-k}u_k$. The state is $(\text{II}_2, \text{II}_2)$.

2. $X = l^1$, $Y = l^2$, A as in problem 1. The state is $(\text{II}_2, \text{III}_2)$.

3. $X = Y = l^2$. $Au_1 = 0$, $Au_k = u_k$ if $k \geqslant 2$. The state is $(\text{III}_3, \text{III}_3)$.

4. $X = Y = l^2$. $Au_1 = 0$, $Au_k = u_{k-1}$, $k \geqslant 2$. The state is $(\text{I}_3, \text{III}_1)$.

5. $X = Y = l^2$. $Au_k = u_{k+1}$. The state is $(\text{III}_1, \text{I}_3)$.

6. $X = Y = l^2$. $Au_1 = 0$, $Au_k = \sum_{i=k-1}^{\infty} 2^{1-i}u_i$. The state is $(\text{II}_3, \text{III}_2)$. Note that $A(u_k - u_{k+1}) = 2^{2-k}u_{k-1}$ if $k \geqslant 2$. Also that $\sum_1^{\infty} 2^{1-k}u_{2k}$ is not in $\mathscr{R}(A)$.

7. $X = Y = l^2$. $Au_k = 2^{1-k}(u_2 + \cdots + u_{k+1})$. The state is $(\text{III}_2, \text{II}_3)$. We can also take $X = Y = (c_0)$ in this case.

8. $X = Y = l^2$. $Au_1 = u_2$, $Au_k = u_{k-1} + u_{k+1}$ if $k \geqslant 2$. The state is $(\text{II}_2, \text{II}_2)$. Note that u_{2k} is in $\mathscr{R}(A)$, but u_{2k-1} is not. But it can be shown that $u_{2k-1} \in \overline{\mathscr{R}(A)}$.

9. $X = Y = (c_0)$. $Au_1 = 2u_1 - u_2$, $Au_k = -u_{k-1} + 2u_k - u_{k+1}$ if $k \geqslant 2$. The state is $(\text{II}_2, \text{II}_2)$.

10. Suppose $A \in [X, Y]$ and $\mathscr{R}(A) = Z \neq Y$. Define $B \in [X, Z]$ by setting $Bx = Ax$ when $x \in X$. Suppose also that Y is complete and $\overline{\mathscr{R}(A)} = Y$. Then we can identify Z' with Y' and B' with A'. The states of A' and B' are the same, but, if the state of A is II_i $(i = 1, 2, 3)$, that of B is I_i. In this way we can produce examples of states $(\text{I}_2, \text{II}_2)$, $(\text{I}_2, \text{III}_2)$, and $(\text{I}_3, \text{III}_2)$ from some of the earlier examples.

11. X the subspace of l^2 generated by $\{u_k\}$, $Y = l^2$. A as in problem 4, but with this change in X. The state is $(\text{II}_3, \text{III}_1)$.

12. X the subspace of (c_0) generated by $\{u_k\}$, Y either X or (c_0) itself. $Au_1 = u_1$, $Au_k = -2u_{k-1} + u_k$ if $k \geqslant 2$. The state is $(\text{II}_2, \text{III}_1)$ if $Y = (c_0)$ and $(\text{I}_2, \text{III}_1)$ if $Y = X$.

13. If Y is a complete space and X is a proper but dense subspace of Y, let $Ax = x$ for $x \in X$. Then $A \in [X, Y]$ and the state of the pair is $(\text{II}_1, \text{I}_1)$.

14. X the subspace of l^1 generated by $\{u_k\}$, Y either X or l^1 itself. $Au_1 = \sum_1^{\infty} 2^{-k}u_k$, $Au_k = u_{k-1}$ if $k \geqslant 2$. The state is $(\text{II}_2, \text{III}_1)$ or $(\text{I}_2, \text{III}_1)$ according to whether Y is l^1 or X.

15. Let Y be any Banach space of infinite dimension, and let H be a Hamel basis for Y with all elements h of H such that $\|h\| \leqslant 1$. If $y \in Y$ and $y = \alpha_1 h_1 + \cdots + \alpha_n h_n$, let $N(y) = \sum_{i=1}^{n} |\alpha_i|$. Let X be the space with the same elements as Y, but with N as a norm. Let $Ay = y$, and consider A as an element of $[X, Y]$. The state is $(\text{I}_2, \text{III}_1)$.

16. If X and Y are linearly homeomorphic and X is norm reflexive, so is Y. Suggestion: Let A be the linear homeomorphism of X onto Y. Determine the state of A'' by using the state diagram. Consider the canonical mappings: J_1 of X onto X'' and J_2 of Y into Y''. Express J_2 in terms of J_1, A'' and A^{-1}.

17. Suppose $A \in [X, X']$. Assume that $\overline{\mathscr{R}(A)} = \overline{\mathscr{R}(A')}$. If X is norm reflexive, show that the only possible states are (I_1, I_1), $(\text{II}_2, \text{II}_2)$, $(\text{III}_3, \text{III}_3)$.

18. Suppose $A \in [X, X']$, that $A = A'J$, and that X is norm reflexive. Then, the states of A and A' are the same, and the only possible states of A are I_1, II_2, III_3.

19. Suppose $A \in [X]$. Let B be a linear homeomorphism of X onto X', and suppose $A = B^{-1}A'B$. Then, the only possible states for A are I_1, II_2, III_3.

4.8 Projections

Operations of a special type, called *projections*, play an important role in the systematic study of linear operators. The notion of a projection is closely related to the concept of a *direct sum of linear manifolds*, and it is with this concept that we begin this section.

Let $M_1, \cdots, M_n (n \geqslant 2)$ be linear manifolds in a linear space X. We say that this set of linear manifolds is linearly independent if no M_i contains a nonzero vector which is in the linear manifold determined by the remaining $n - 1$ linear manifolds. An equivalent condition is that if $x_i \in M_i$ and $\alpha_1 x_1 + \cdots + \alpha_n x_n = 0$, then $x_k = 0$ if $\alpha_k \neq 0$. The linear manifold generated by the elements of $M_1 \cup \cdots \cup M_n$ is denoted by $M_1 \oplus \cdots \oplus M_n$ and called the *direct sum* of $M_1, \cdots, M_n$. Elements x of the direct sum are representable uniquely in the form $x = x_1 + \cdots + x_n$, with $x_i \in M_i$.

Whenever we use the $\oplus$ notation, it is to be understood implicitly that the manifolds in question are linearly independent.

If $X = M_1 \oplus \cdots \oplus M_n$, the representation $x = x_1 + \cdots + x_n$, with $x_i \in M_i$, determines operators $P_1, \cdots, P_n$, defined by $P_i x = x_i$. These operators are linear, and they satisfy the relations

$$\text{(4.8-1)} \qquad \begin{cases} P_i P_j = 0 \quad \text{if } i \neq j, \quad P_i^2 = P_i, \\ I = P_1 + \cdots + P_n. \end{cases}$$

Conversely, if the linear operators $P_1, \cdots, P_n$ are given, with domains X

and ranges $M_1, \cdots, M_n$ in X, and if they satisfy the conditions (4.8–1), it follows that $X = M_1 \oplus \cdots \oplus M_n$.

Definition. A linear operator P with domain X and range in X is called a *projection* (of X) if $P^2 = P$. If M is the range of P, then P is called a projection of X onto M.

It is easily seen that, if P is a projection of X, then

$$X = \mathscr{R}(P) \oplus \mathscr{N}(P). \tag{4.8–2}$$

The direct sum representation of an element is

$$x = Px + (x - Px).$$

The operator $Q = I - P$ is also a projection; its range is $\mathscr{N}(P)$. Elements x of $\mathscr{R}(P)$ are characterized by the fact that $Px = x$.

When $X = M_1 \oplus M_2$, we say that M_2 is a complement of (or is complementary to) M_1. Thus, (4.8–2) says that, if P is a projection, $\mathscr{N}(P)$ is complementary to $\mathscr{R}(P)$. There can be more than one complement of a given linear manifold.

As for the existence of complements of a given linear manifold, we have the following result:

Theorem 4.8–A. *If M is a linear manifold in the linear space X, there exists a projection P of X whose range is M.*

PROOF. The cases $M = (0)$ or X are obvious, so we set them aside. Let f be the linear operator with domain and range M, defined by $f(x) = x$. Take $Y = M$ in Theorem 1.71–A. Then the F of Theorem 1.71–A will serve as P for our present purpose.

Thus far nothing has been said about topology in connection with projections.

Theorem 4.8–B. *If X is a topological linear space and if P is a projection of X which is closed, the range of P is closed.*

PROOF. Suppose $y \in \overline{\mathscr{R}(P)}$. If V is any neighborhood of y, it contains a point of $\mathscr{R}(P)$, say $z = Px$. Then $Pz = z$, and (z, z) is a point of the graph of P. But also, $(z, z) \in V \times V$, which is a neighborhood of (y, y) in the product space $X \times X$. Hence, since V is arbitrary and the graph of P is closed, (y, y) is in the graph. That is, $Py = y$, and so $y \in \mathscr{R}(P)$. Hence $\mathscr{R}(P)$ is closed.

Theorem 4.8–C. *Suppose X is a topological linear T_1-space. Suppose P is a continuous projection of X. Then $\mathscr{R}(P)$ is closed.*

PROOF. $\mathscr{R}(P)$ is the null space of the continuous operator $I - P$. Since (0) is closed, so is the inverse image $\{x : x - Px = 0\}$.

Theorem 4.8–D. *Let X be a complete metric linear space. Let P be a projection of X such that both $\mathscr{R}(P)$ and $\mathscr{N}(P)$ are closed. Then P is continuous.*

PROOF. Because of Theorem 4.2–I it suffices to prove that P is closed. Suppose $x_n \to x$ and $Px_n \to y$. Then $x_n - Px_n \to x - y$. Since $Px_n \in \mathscr{R}(P)$ and $x_n - Px_n \in \mathscr{N}(P)$, it follows that $y \in \mathscr{R}(P)$ and $x - y \in \mathscr{N}(P)$. Then $Px - Py = 0$, or $Px = Py = y$. Thus P is closed.

If X is a Banach space and M is a closed subspace of X, there may be no continuous projection of X whose range is M. In view of (4.8–2) and Theorem 4.8–D, this is the same as saying that there may be no *closed* linear manifold which is complementary to M. Examples have been given, even in reflexive spaces, by F. J. Murray, **1** and by A. Sobczyk, **1**. It can also occur, even in Hilbert space, that $M_1 \oplus M_2$ is not closed, even though M_1 and M_2 are closed and linearly independent subspaces. In this case the associated projections of $M_1 \oplus M_2$ onto M_1 and M_2 are not continuous. For an example, see Stone, **1**, pages 21–22.

For inner-product spaces the notion of direct sum of linear manifolds is of particular importance in connection with the concept of mutually orthogonal linear manifolds.

Definition. Two linear manifolds M, N in the inner-product space X are said to be orthogonal if $(x, y) = 0$ whenever $x \in M$ and $y \in N$. A family of linear manifolds is called an orthogonal family if each pair of distinct manifolds from the family are orthogonal.

If $M_1, \cdots, M_n$ is an orthogonal family of linear manifolds, the manifolds are linearly independent. For, if $\alpha_1 x_1 + \cdots + \alpha_n x_n = 0$, with $x_k \in M_k$, the orthogonality shows that $0 = (\alpha_1 x_1 + \cdots + \alpha_n x_n, x_k) = \alpha_k \|x_k\|^2$, and hence $x_k = 0$ if $\alpha_k \neq 0$.

A direct sum of orthogonal closed linear manifolds in a complete space is closed. We state this formally.

Theorem 4.8–E. *If X is a complete inner-product space and if M, N are orthogonal closed linear manifolds, $M \oplus N$ is closed.*

PROOF. If $x \in M$ and $y \in N$, an easy calculation shows that $\|x + y\|^2 = \|x\|^2 + \|y\|^2$. Now suppose that $z_n \in M \oplus N$ and that $z_n \to z$. We can write $z_n = x_n + y_n$, $x_n \in M$, $y_n \in N$. Then

$$\|z_n - z_m\|^2 = \|x_n - x_m\|^2 + \|y_n - y_m\|^2.$$

Since $\{z_n\}$ is a Cauchy sequence, so are $\{x_n\}$ and $\{y_n\}$. If $x_n \to x$ and $y_n \to y$, we have $z = x + y$. But $x \in M$, $y \in N$, since M and N are closed. Hence $z \in M \oplus N$, and the proof is complete.

PROBLEMS

1. Suppose X is a Banach space, $M_1, \cdots, M_n$ are closed and linearly independent subspaces of X, and $M = M_1 \oplus \cdots \oplus M_n$. Suppose that the projections $P_1, \cdots, P_n$ of M onto $M_1, \cdots, M_n$ respectively, associated with the direct sum, are continuous. Then M is closed.

2. If X is a complete metric linear space, if $M = M_1 \oplus M_2$, where M_1 and M_2 are closed and linearly independent subspaces, and if M is closed, the associated projections of M onto M_1 and M_2 are continuous. This does not generalize for more than two M_i's.

3. If X is a Banach space and M_1, M_2 are closed and linearly independent subspaces of X, then $M_1 \oplus M_2$ is closed if and only if there exists a $d > 0$ such that $\|x_1 - x_2\| \geqslant d$ whenever $x_1 \in M_1$, $x_2 \in M_2$, and $\|x_1\| = \|x_2\| = 1$.

4. Let X be a normed linear space. Let P_1 and P_2 be continuous projections of X such that $I = P_1 + P_2$. Let M_k be the range of P_k, so that $X = M_1 \oplus M_2$. Then P_k' is a continuous projection of X'. The range of P_1' is $M_2{}^0$, and the range of P_2' is $M_1{}^0$. Finally, $X' = M_1{}^0 \oplus M_2{}^0$.

5. If P is a projection on any linear space, show that $(\lambda - P)^{-1} = \lambda^{-1}I + \lambda^{-1}(\lambda - 1)^{-1}P$ if $\lambda(\lambda - 1) \neq 0$.

4.81 Continuous Linear Functionals on a Hilbert Space

We have seen, in Theorem 4.32–A, that every continuous linear functional on l^2 is representable in the form

$$x'(x) = \sum_1^\infty \alpha_n \xi_n, \qquad x = \{\xi_n\} \in l^2,$$

where

$$\|x'\| = \left(\sum_1^\infty |\alpha_n|^2\right)^{\frac{1}{2}} < \infty.$$

This result can be restated as follows, by defining $y = \{\bar{\alpha}_n\}$ as an element of l^2: to each $x' \in (l^2)'$ corresponds a unique $y \in l^2$ such that $x'(x) = (x, y)$ for each $x \in l^2$. In this form the result is valid for an arbitrary Hilbert space, or, since it is true for finite-dimensional as well as for infinite-dimensional spaces, it is valid for all complete inner-product spaces. To prove this we need two preliminary theorems. In each of the three following theorems it is assumed that X is a complete inner-product space.

Theorem 4.81–A. *Let S be a nonempty closed convex set in X, and suppose $d = \inf_{x \in S} \|x\|$. Then, there exists a unique $x \in S$ such that $\|x\| = d$.*

PROOF. From the way in which the norm is defined in terms of the inner product, it follows that

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2) \tag{4.81–1}$$

for all pairs x, y in X. Now suppose that $x_n \in S$ and $\|x_n\| \to d$. Since S is convex, $\frac{1}{2}(x_n + x_m) \in S$, and so

$$\left\|\frac{x_n + x_m}{2}\right\| \geqslant d.$$

From this inequality and (4.81–1) we have

$$\begin{aligned}\|x_n - x_m\|^2 &= 2(\|x_n\|^2 + \|x_m\|^2) - \|x_n + x_m\|^2 \\ &\leqslant 2(\|x_n\|^2 + \|x_m\|^2) - 4d^2,\end{aligned}$$

whence, since $\|x_n\|^2 \to d^2$, it appears that $\{x_n\}$ is a Cauchy sequence. If $x_n \to x$, it follows that $x \in S$ and $\|x\| = d$. To see that x is unique, suppose x and y are in S and that $\|x\| = \|y\| = d$. Then, since S is convex,

$$d \leqslant \left\|\frac{x + y}{2}\right\| \leqslant \tfrac{1}{2}\|x\| + \tfrac{1}{2}\|y\| = d.$$

It then follows by (4.81–1) that

$$4d^2 + \|x - y\|^2 = 4d^2.$$

whence $x = y$.

Theorem 4.81–B. *Let H be the closed hyperplane $\{x : y'(x) = 1\}$ where y' is a fixed nonzero element of X'. Then, there is a unique $z \in H$ such that $(x, z) = 0$ whenever $y'(x) = 0$.*

PROOF. H is nonempty, closed, and convex. Hence, by Theorem 4.81–A, there exists $z \in H$ such that $\|z\| = \inf_{x \in H} \|x\|$. Now choose any x such that $x \neq 0$ and $y'(x) = 0$. We wish to prove that $(x, z) = 0$. Define

$$z_1 = \frac{(z, x)}{\|x\|^2} x, \qquad z_2 = z - z_1. \tag{4.81–2}$$

Then $y'(z_1) = 0$, and so $y'(z_2) = y'(z) = 1$ (since $z \in H$). Hence $z_2 \in H$. It follows that $\|z_2\| \geqslant \|z\|$. But, from (4.81–2) it follows that $(z_1, z_2) = 0$ and hence that

$$\|z\|^2 = (z_1 + z_2, z_1 + z_2) = \|z_1\|^2 + \|z_2\|^2, \tag{4.81–3}$$

whence $\|z_2\| \leqslant \|z\|$. Therefore $\|z_2\| = \|z\|$, and consequently, by (4.81–3), $z_1 = 0$. Since $x \neq 0$, this implies $(x, z) = 0$. To see that z is unique, suppose that w also has the properties asserted for z. Then $y'(w - z) = y'(w) - y'(z) = 1 - 1 = 0$, and hence $(w - z, z) = (w - z, w) = 0$. But then $\|w - z\|^2 = 0$, or $w = z$.

We now come to the main representation theorem.

Theorem 4.81–C. *To each $y' \in X'$ corresponds a unique $y \in X$ such that $y'(x) = (x, y)$ for every x.*

PROOF. It is clear that there cannot be more than one such y, for if y_1 and y_2 meet the requirement, it follows that $(x, y_1 - y_2) = 0$ for every x, so that, in particular, $\|y_1 - y_2\|^2 = 0$ and $y_1 = y_2$. If $y' = 0$ we can take $y = 0$, so we assume $y' \neq 0$. Choose z as in Theorem 4.81–B. Now, (x, z) is a linear functional of x; hence, by Theorem 3.5–C, there is some scalar α such that $(x, z) = \alpha y'(x)$ for every x. Putting $x = z$, we obtain $\|z\|^2 = \alpha y'(z) = \alpha$. Hence $y'(x) = (x, z)/\|z\|^2$ for every x. We now let $y = z/\|z\|^2$ and the proof is complete.

PROBLEMS

1. If y', y, and z are related as in the proof of Theorem 4.81–C, show that $\|y'\| = \|y\| = 1/\|z\|$. Thus the distance from 0 to the set $\{x : y'(x) = 1\}$ is $1/\|y'\|$.

2. If X is a complete inner-product space, X and X' are congruent. For the complex case use the result at the end of § 3.2.

3. A Hilbert space is norm reflexive.

4. Let M be a closed linear manifold in the complete inner-product space X. If $x \in X$, define Qx as the unique element of $N = x + M$ such that $\|Qx\| = \inf_{y \in N} \|y\|$ (use Theorem 4.81–A). From $\|Qx\|^2 \leqslant \|Qx + \alpha m\|^2$ if $m \in M$, show that Qx is orthogonal to M [choose $\alpha = -(Qx, m)/\|m\|^2$]. Note that $x = (x - Qx) + Qx$ and that $x - Qx \in M$. The element $x - Qx$ is called the *orthogonal projection of x onto M*. Observe that $Qx = 0$ if $x \in M$.

5. Use the result of problem 4 to give an alternative proof of Theorem 4.81–B. Take $M = \{x : y'(x) = 0\}$.

4.82 Orthogonal Complements

In this section X denotes an inner-product space. Completeness is not assumed except as stated explicitly.

Definition. If S is a nonempty subset of X, the set of all x such that $(x, y) = 0$ if $y \in S$ is called the *orthogonal complement* of S, and is denoted by $S^{\perp}$.

A number of simple facts are at once evident:

$$(0)^{\perp} = X, \quad X^{\perp} = (0),$$
$$S_1 \subset S_2 \text{ implies } S_2^{\perp} \subset S_1^{\perp},$$
$$x \in S \cap S^{\perp} \text{ implies } x = 0.$$

The set $S^\perp$ is a closed linear manifold. If M is the closed linear manifold generated by S, then $M^\perp = S^\perp$.

We denote $(S^\perp)^\perp$ by $S^{\perp\perp}$. It is obvious that $S \subset S^{\perp\perp}$.

The main theorem about orthogonal complements is the following:

Theorem 4.82–A. *If X is complete and if M is a closed subspace of X, then $X = M \oplus M^\perp$.*

PROOF. Choose any $y \in X$. The inner product (x, y), as x varies over M, defines an element of M'. Since M can be regarded as a complete inner-product space, Theorem 4.81–C asserts the existence of an element $y_1 \in M$ such that $(x, y) = (x, y_1)$ if $x \in M$. Let $y_2 = y - y_1$. Then $y_2 \in M^\perp$, for $(x, y_2) = (x, y) - (x, y_1) = 0$ if $x \in M$. Writing $y = y_1 + y_2$, we have the proof of the theorem.

If we write $Py = y_1$ in the foregoing notation, P is a projection with range M. It is called *the orthogonal projection of X onto M*, because of the fact that $x - Px \in M^\perp$. We note that a different proof of Theorem 4.82–A is indicated in problem 4 of § 4.81.

PROBLEM

Suppose X is a complete inner-product space. If $y' \in X'$ is represented by $y'(x) = (x, y)$, let $Ey = y'$, so that E is an operator which maps X onto X'. Show that annihilators (as in § 4.6) and orthogonal complements are related as follows: If $S \subset X$, then

$$S^0 = E(S^\perp), \qquad S^\perp = E^{-1}(S^0),$$

and ${}^0(S^0) = S^{\perp\perp}$.

4.83 Dirichlet's Principle

The subject matter of this section illustrates the usefulness, in potential theory, of the notion of the direct sum of two orthogonal subspaces of an inner-product space.

Let R be a bounded open set with boundary $B(R)$, in Euclidean space of three dimensions. We suppose, furthermore, that $B(R)$ is smooth enough to enable us to apply the divergence theorem to vector fields which are continuously differentiable in $\bar{R}$. In particular, then, with adequate continuity and differentiability we have the identity of Green

$$\int_R g\nabla^2 h\, dV + \int_R \nabla g \cdot \nabla h\, dV = \int_{B(R)} g \frac{\partial h}{\partial n}\, dS, \tag{4.83–1}$$

where g and h are real-valued functions defined in R, ∇^2 denotes the Laplacian, ∇ the gradient, and n the outward normal to $B(R)$.

We define a real inner-product space $\mathscr{F}$ as follows: The elements f of $\mathscr{F}$ are the real-valued functions which are twice continuously differentiable in $\bar{R}$. The inner product is

$$(f_1, f_2) = \int_{B(R)} f_1 f_2 \, dS + \int_R \nabla f_1 \cdot \nabla f_2 \, dV. \tag{4.83-2}$$

The space $\mathscr{F}$ is not complete. Let $\mathscr{G}$ be the subspace of those elements $g \in \mathscr{F}$ which vanish on $B(R)$, and let $\mathscr{H}$ be the subspace of those elements $h \in \mathscr{F}$ which are harmonic (i.e., $\nabla^2 h = 0$) in R. From (4.83–1) and (4.83–2) we see that $(g, h) = 0$ if $g \in \mathscr{G}$ and $h \in \mathscr{H}$; that is, $\mathscr{G}$ and $\mathscr{H}$ are orthogonal.

Now consider any $h \in \mathscr{H}$ and any $f \in \mathscr{F}$ such that $f = h$ on $B(R)$. Let $g = f - h$. Then $g \in \mathscr{G}$. Since $f = g + h$, the orthogonality shows that

$$\|f\|^2 = \|g\|^2 + \|h\|^2,$$

and hence $\|h\| \leqslant \|f\|$. Thus, among all elements of $\mathscr{F}$ which are equal to the given h on $B(R)$, h has the smallest norm. This characterization of h by a minimal property is what is known as *Dirichlet's principle* (though originally the principle was not stated in the language of norms).

In the early work in the calculus of variations, attempts were made to use Dirichlet's principle to prove the existence of a solution of the Dirichlet problem. Let us see how we might attempt to give an existence proof by linear-space methods. Suppose f_0 is a given element of $\mathscr{F}$. Denote by K the set of elements of $\mathscr{F}$ which coincide with f_0 on $B(R)$. This set is precisely the set $f_0 + \mathscr{G}$. Since $\mathscr{G}$ is a closed linear manifold in $\mathscr{F}$, K is a closed-linear variety in $\mathscr{F}$; in particular, K is convex. If f_0 does not vanish identically on $B(R)$, the distance d from 0 to K is positive, and there exists a sequence $\{f_n\}$ in K such that $\|f_n\| \to d$. Just as in the proof of Theorem 4.81–A we see that $\{f_n\}$ is a Cauchy sequence. But, since $\mathscr{F}$ is not complete, we cannot be certain that the sequence has a limit in $\mathscr{F}$. (It is precisely here that this method fails in its attempt to provide an existence proof for the Dirichlet problem.) If a limit h *does* exist in $\mathscr{F}$, then $h \in K$. In this case, the fact that h is the element of K with minimal norm allows us to conclude, by the standard reasoning of the calculus of variations, that h satisfies Laplace's equation, which is the Euler equation for this situation. Thus we can say: If $f_0 \in \mathscr{F}$, the Dirichlet problem for the region R, with the values of f_0 on $B(R)$ as assigned boundary values, is solvable if and only if there is a point of $f_0 + \mathscr{G}$ at minimal distance from 0. This "point" is then the solution of the Dirichlet problem.

It is interesting to observe that, if the Dirichlet problem is solvable, in the sense just described, for each $f \in \mathscr{F}$, we can write

$$\mathscr{F} = \mathscr{G} \oplus \mathscr{H}.$$

Hence, if $f = g + h$, where $g \in \mathscr{G}$ and $h \in \mathscr{H}$, then h is the solution of the Dirichlet problem with boundary values of f. We can describe h as the orthogonal projection of $\mathscr{F}$ on $\mathscr{H}$.

This concept of orthogonal projections in connection with the Dirichlet problem became very well known in the 1930's and was utilized in various ways by research workers in the calculus of variations. It led naturally to various schemes for overcoming the difficulties arising from an incomplete space. One line of development was that in which the space was taken to be complete and the basic Dirichlet problem was modified by generalizing the sense in which a function takes on prescribed values at the boundary of a region. We present a brief exposition of a paper by Weyl (**1**, especially pages 411–414). We now let R be any open set in Euclidean 3-space. We consider vector fields F of class $\mathscr{L}^2$ in R and form a Hilbert space $\mathscr{F}$ with the inner product

$$(F_1, F_2) = \int_R F_1 \cdot F_2 \, dV,$$

by the usual formation of equivalence classes. We shall be interested in several subspaces of $\mathscr{F}$. First we consider the vector-valued functions defined in R, each of which has the value zero outside of some compact subset of R and has continuous first derivatives in R. If G is such a function, its curl $\nabla \times G$ determines an element of $\mathscr{F}$; we denote by $\mathscr{C}$ the closed linear manifold in $\mathscr{F}$ generated by all such $\nabla \times G$. Next we consider real-valued functions defined and continuously differentiable in R, each of which vanishes outside of some compact subset of R. The gradients ∇v of functions v of this kind generate a certain closed linear manifold in $\mathscr{F}$, which we denote by $\mathscr{G}$. The elements of $\mathscr{F}$ which are orthogonal to $\mathscr{C}$ are called *irrotational*; those which are orthogonal to $\mathscr{G}$ are called *solenoidal.* These are extensions of the ordinary meanings of these terms. We write $\mathscr{F}_i$ for the class of irrotational elements and $\mathscr{F}_s$ for the class of solenoidal elements. Note, in particular, that $\mathscr{G} \subset \mathscr{F}_i$ and $\mathscr{C} \subset \mathscr{F}_s$. That is, $(\nabla v, \nabla \times G) = 0$ for any v and G of the sort mentioned above. This would be obvious from the relation

$$\nabla \cdot (G \times \nabla v) = \nabla v \cdot (\nabla \times G) - G \cdot (\nabla \times \nabla v)$$

if v were twice continuously differentiable. With the weaker conditions here placed on v a deeper examination of the situation is required. The discussion given by Weyl attends to this matter.

By Theorem 4.82–A we have $\mathscr{F} = \mathscr{G} \oplus \mathscr{F}_s$, since $\mathscr{F}_s = \mathscr{G}^{\perp}$. Now consider the elements of $\mathscr{F}_i$ which are orthogonal to $\mathscr{G}$. These are precisely the elements of $\mathscr{F}_i \cap \mathscr{F}_s$. Hence

$$\mathscr{F}_i = \mathscr{G} \oplus (\mathscr{F}_i \cap \mathscr{F}_s).$$

The relation of all this to the Dirichlet problem (as well as to other problems of potential theory) depends on the following fundamental fact, which is proved in the reference already cited: If $F \in \mathscr{F}_i \cap \mathscr{F}_s$, then F has derivatives of all orders in R, and the divergence and curl of F are both identically zero in R. It follows that the components of F are harmonic functions. Also, F is the gradient of a harmonic function; however, the latter function may be multiple valued, owing to the fact that R need not be simply connected. Suppose now that we start with a given scalar function f such that $F = \nabla f$ is in $\mathscr{F}$. Then $F \in \mathscr{F}_i$, and we can write $F = G + H$, where $H = \nabla h$, h being harmonic, $G = \lim_{n\to\infty} \nabla g_n$, and g_n vanishes outside of some compact subset of R. It can be shown that this behavior of g_n implies that, in a generalized sense, $f - h$ is zero at the boundary of R. Thus h is the solution of a generalized Dirichlet problem.

4.9 Adjoint Operators

Throughout this section X and Y will denote arbitrary complete inner-product spaces. We use the same symbol (parentheses) for the inner product in both spaces. If $A \in [X, Y]$, we have defined A' so that $A' \in [Y', X']$. However, because of Theorem 4.81–C, it is possible for us to identify X' with X and Y' with Y. In so doing, we are led to an operator belonging to $[Y, X]$ which it is natural to consider in place of A'. Suppose y_0 is fixed in Y. Let $x'(x) = (Ax, y_0)$. Then $x' \in X'$, and hence (by Theorem 4.81–C) there is a unique $x_0 \in X$ such that $x'(x) = (x, x_0)$. We write $x_0 = A^*y_0$, thus defining an operator A^* on Y into X. The definition of A^* is fully expressed by the equation

$$(Ax, y) = (x, A^*y), \qquad x \in X,\ y \in Y. \tag{4.9–1}$$

The operator A^* is called the *adjoint* of A.

It is easy to see that A^* is linear. It is also continuous, and

$$\|A^*\| = \|A\|. \tag{4.9–2}$$

For hints as to the proof of this, see problem 1.

The following formal properties of adjoints are easily verified:

$$(A + B)^* = A^* + B^*, \qquad (\alpha A)^* = \bar{\alpha}A^*, \qquad 0^* = 0.$$

If I is the identity operator on X to itself, $I^* = I$. If A^{-1} exists and belongs to $[Y, X]$, then $(A^*)^{-1}$ exists, belongs to $[X, Y]$, and

$$(A^*)^{-1} = (A^{-1})^*. \tag{4.9–3}$$

If X, Y, Z are all complete inner-product spaces and if $A \in [Y, Z]$ and $B \in [X, Y]$, then $AB \in [X, Z]$ and

$$(AB)^* = B^*A^*. \tag{4.9–4}$$

We write $(A^*)^* = A^{**}$. We often use the relation

$$A^{**} = A, \tag{4.9–5}$$

which follows easily from (4.9–1).

The relations between annihilators, ranges, and null manifolds, as set forth in § 4.6, have counterparts in relations between orthogonal complements, ranges, and null manifolds. The situation is simpler than that of § 4.6, owing to the identification of X' with X. A complete inner-product space is reflexive; and there are no nonsaturated subspaces of X'. We list the following results, leaving verification to the reader.

$$\{\overline{\mathscr{R}(A)}\}^\perp = \mathscr{N}(A^*); \qquad \overline{\mathscr{R}(A)} = \mathscr{N}(A^*)^\perp. \tag{4.9–6}$$

$$\{\overline{\mathscr{R}(A^*)}\}^\perp = \mathscr{N}(A); \qquad \overline{\mathscr{R}(A^*)} = \mathscr{N}(A)^\perp. \tag{4.9–7}$$

Next we state two simple but useful theorems.

Theorem 4.9–A. *$\mathscr{N}(A^*) = \mathscr{N}(AA^*)$, and $\overline{\mathscr{R}(A)} = \overline{\mathscr{R}(AA^*)}$.*

PROOF. Obviously $\mathscr{N}(A^*) \subset \mathscr{N}(AA^*)$. If $AA^*y = 0$, we have $(A^*y, A^*y) = (AA^*y, y) = 0$, and so $A^*y = 0$. Hence $\mathscr{N}(AA^*) \subset \mathscr{N}(A^*)$. The first part of the theorem is now proved. Next, we observe that $(AA^*)^* = A^{**}A^* = AA^*$. Hence, by (4.9–6) and the first part of our theorem, $\overline{\mathscr{R}(AA^*)} = \mathscr{N}(AA^*)^\perp = \mathscr{N}(A^*)^\perp = \overline{\mathscr{R}(A)}$. This finishes the proof.

Theorem 4.9–B. $\|A^*A\| = \|A\|^2 = \|AA^*\|$.

PROOF. $\|A^*Ax\| \leqslant \|A^*\|\,\|Ax\| \leqslant \|A^*\|\,\|A\|\,\|x\| = \|A\|^2\|x\|$, by (4.9–2). Therefore $\|A^*A\| \leqslant \|A\|^2$. But also, $\|Ax\|^2 = (Ax, Ax) = (x, A^*Ax) \leqslant \|x\|\,\|A^*Ax\| \leqslant \|A^*A\|\,\|x\|^2$, whence $\|A\|^2 \leqslant \|A^*A\|$. Thus $\|A^*A\| = \|A\|^2$.

We apply this result with A^* in place of A, obtaining $\|A^{**}A^*\| = \|AA^*\| = \|A^*\|^2 = \|A\|^2$. Thus all is proved.

Let us now recall from § 4.71 the terminology about the nine possible *states* of a bounded linear operator. Since we are now considering A^* instead of A', it is important to know that the state of A^* is always the same as the state of A'. That this is so follows from the very close relationship between A^* and A'; we leave verification to the reader. As a consequence, A^* may replace A' in the state diagram in § 4.71. Since X and Y are reflexive, there are just seven possible states for the pair (A, A^*).

We say that A is *normal* if $AA^* = A^*A$ and *self-adjoint* if $A^* = A$. Clearly a self-adjoint operator is normal, but the converse does not always hold. For a normal operator the number of possible states is severely limited.

Theorem 4.9–C. *For a normal operator A the only possible states are* I_1, II_2, *and* III_3. *The state of A^* is the same as that of A.*

PROOF. By Theorem 4.9–A, applied to A^*, we have $\overline{\mathscr{R}(A^*)} = \overline{\mathscr{R}(A^*A)}$. Since A is normal, we then have $\overline{\mathscr{R}(A)} = \overline{\mathscr{R}(A^*)}$. It follows from this that neither of the operators A, A^* can be in a III state unless the other is also. If we now examine the state diagram (with A^* in place of A'), we see that four of the seven possible states for the pair (A, A^*) are ruled out, leaving only three states for the pair as indicated in the theorem.

The results of this section are used in Chapter 6.

PROBLEMS

1. Put $x = A^*y$ in (4.9–1) to prove $\|A^*\| \leqslant \|A\|$. Put $y = Ax$ to prove the reverse inequality.

2. Verify (4.9–3), (4.9–4), and (4.9–5).

3. Prove (4.9–6). Then use (4.9–5) to get (4.9–7).

4. Let E be the operator defined in the problem at the end of § 4.82, and let F be the corresponding operator for Y. Show that $A^* = E^{-1}A'F$. Use this to prove that the states of A^* and A' are the same. Notice that E and F are norm-preserving operators.

5. $\mathscr{R}(A) = Y$ if and only if A^* has a continuous inverse. $\mathscr{R}(A^*) = X$ if and only if A has a continuous inverse.

6. Let X be a Banach space, and suppose $A \in [X]$. Then, there exists $B \in [X]$ such that $BA = I$ if and only if (*a*) A has a continuous inverse, and (*b*) there exists a continuous projection P of X onto $\mathscr{R}(A)$. If X is a complete inner-product space, (*b*) is satisfied whenever (*a*) is. Give details of the argument, noting that, if P is known, we can take $B = A^{-1}P$ and that, if B is known, we can take $P = AB$.

SUGGESTED READING

Achieser and Glasmann, **1**, Chapters 2, 3, 4.
Banach, **1**, Chapters 3–11.
Cooke, **1**, Chapters 2, 3.
Dunford and Schwartz, **1**, Chapters 2, 4, 5, 6.
Friedman, **1**, Chapter 1.
Halmos, **3**, parts of Chapters 1, 2.
Hellinger and Toeplitz, **1**, Parts I, II.
Hille, **1**, Chapters 2, 3.
Julia, **1**, Chapters 4, 5.
Ljusternik and Sobolew, **1**, Chapters 2, 3.
Loomis, **1**, Chapter 2.
Riesz and Nagy, **1**, parts of Chapters 4, 5.
Stone, **1**, Chapters 2, 3.
Wintner, **1**, Chapter 3.
Zaanen, **1**, Chapters 6, 7, 9, 10.

5

SPECTRAL ANALYSIS OF LINEAR OPERATORS

5.0 Introduction

In this preliminary section we shall indicate the main trend of the ideas to be developed in this chapter. We consider a normed linear space X and a linear operator T whose domain $\mathscr{D}(T)$ and range $\mathscr{R}(T)$ lie in X. Then we consider the operator $\lambda I - T$, where λ is a scalar parameter and I is the identity operator. For convenience we usually suppress the I and write $\lambda - T$ in place of $\lambda I - T$.

Definition. If λ is such that the range of $\lambda - T$ is dense in X and if $\lambda - T$ has a continuous inverse, we say that λ is in the *resolvent set* of T; this set of values of λ is denoted by $\rho(T)$. All scalar values of λ not in $\rho(T)$ comprise the set called the *spectrum* of T; it is denoted by $\sigma(T)$.

In a broad sense, spectral theory, or spectral analysis, of linear operators is the systematic study of the relations between the operators T, $(\lambda - T)^{-1}$, the sets $\rho(T)$, $\sigma(T)$, and various other operators and linear manifolds which enter the picture naturally.

If X is of finite dimension n and if $\mathscr{D}(T) = X$, we can represent T by an $n \times n$ matrix, say (t_{ij}); then $\lambda - T$ is also represented by a matrix, and $\sigma(T)$ is composed of those scalars λ which are the roots of the equation

$$\begin{vmatrix} \lambda - t_{11} & -t_{12} & \cdots & -t_{1n} \\ -t_{21} & \lambda - t_{22} & \cdots & -t_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ -t_{n1} & \lambda - t_{n2} & \cdots & \lambda - t_{nn} \end{vmatrix} = 0.$$

The determinant here is a polynomial of degree n in λ. Hence, if the scalars associated with X are complex, then $\sigma(T)$ contains at least one point, and it may contain as many as, but not more than, n distinct points. If the scalar field is the real field, however, it may be that $\sigma(T)$ is empty. In spectral theory we usually deal with complex scalars, for by so doing we get a richer theory in which we can utilize much of the machinery of the classical theory of functions of a complex variable.

For nearly all of the extensive developments of spectral theory we assume that X is complete and that T is closed. In this case, if $\lambda \in \rho(T)$, the fact that $\lambda - T$ has a continuous inverse implies (see Theorem 4.2–E) that $\mathscr{R}(\lambda - T)$ is closed; since it is dense in X, it is all of X. Hence in this case, when $\lambda \in \rho(T)$ the operator $(\lambda - T)^{-1}$ belongs to $[X]$. It turns out that $\rho(T)$ is an open set in the space of scalars and that $(\lambda - T)^{-1}$ is developable as a power series in $\lambda - \lambda_0$ about each point $\lambda_0 \in \rho(T)$. This power series has coefficients in $[X]$, and it converges in the norm topology of the space of operators. Thus we can regard $(\lambda - T)^{-1}$ as an analytic function defined on $\rho(T)$, with values in $[X]$. This function is called the *resolvent* of T. Any point of $\sigma(T)$ which is a point of accumulation of $\rho(T)$ may be regarded as a "singularity" of the resolvent. If the scalar field is complex, the relation between $(\lambda - T)^{-1}$ and the spectrum can be investigated with the aid of contour integrals in the complex plane.

In the finite-dimensional case, for instance, the spectrum consists of one or more isolated points each of which is, in a certain definite sense, a *pole* of $(\lambda - T)^{-1}$. By using the Laurent expansion of $(\lambda - T)^{-1}$ about each pole, we find certain projections, with the aid of which we express X as a direct sum of linearly independent subspaces, say $X = M_1 \oplus \cdots \oplus M_k$. These subspaces are invariant under T; that is, $T(M_i) \subset M_i$. Part of the interest of this decomposition of X as a direct sum lies in the fact that, when we restrict attention to T as it operates merely on the subspace M_i, we find that its spectrum consists of a single point. If we choose a suitable basis for M_i, the matrix representation of T takes on an especially simple form. In the general case (X not finite dimensional) matters are usually much more complicated, but the pattern of developments in the finite-dimensional case still indicates a certain direction for investigations.

Spectral theory includes the study of eigenvalue problems. If λ is a scalar such that $\lambda - T$ has no inverse, this means that there is at least one nonzero vector x such that $Tx = \lambda x$. In this case λ is called an *eigenvalue* of T, and x is a corresponding *eigenvector*. The null manifold of $\lambda - T$ is called the *eigenmanifold* corresponding to λ. An eigenvalue of T is a point in the spectrum of T. In general, however, $\sigma(T)$ will contain points which are not eigenvalues.

Certain parts of the classical theory of integral equations can be developed within the framework of abstract spectral theory. A Fredholm-type integral operator, acting in the space $C[a, b]$, has certain special properties which can be described in terms which are meaningful in any Banach space, and it turns out that these properties furnish a basis for a general theory (the theory of compact operators) which includes important parts of the theory of Fredholm integral equations of the second kind.

There are certain special features of great interest in the spectral analysis of linear operators in inner-product spaces; this is especially so for Hilbert spaces. The most highly developed spectral theory is that for self-adjoint operators, which are in a certain sense like operators T in $l^2(n)$ which are represented by matrices (t_{ij}) with Hermitian symmetry: $t_{ji} = \bar{t}_{ij}$. For the finite-dimensional case of the space $l^2(n)$ the dominant fact about such operators is that there is a certain orthonormal basis for the space such that the matrix representation of the operator takes the form

$$\begin{Vmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{Vmatrix},$$

where $\lambda_1, \cdots, \lambda_n$ are the points (not necessarily all distinct) forming the spectrum of the operator. The basis vectors in this case are eigenvectors, and $l^2(n)$ is the direct sum of eigenmanifolds corresponding to distinct eigenvalues. The appropriate generalization of these things for operators in Hilbert space is taken up in Chapter 6.

The main theorems on the spectrum and the resolvent of a linear operator are 5.1–B, 5.1–C, 5.2–A, 5.2–B and 5.2–E. The principal theorems in the spectral theory of compact operators are 5.5–F, 5.5–G, 5.5–H, and 5.5–I. Section 5.5 concludes with a summary of applications of the theory of compact operators to the study of integral equations of the second kind. In the latter part of the chapter, where the spectral analysis is carried on by means of the operational calculus founded on contour integrals, the principal theorems are 5.6–A, 5.7–A, 5.71–A, 5.8–A, and 5.8–B.

5.1 The Resolvent Operator

Once and for all we assume, in this chapter, that X is a normed linear space which contains some nonzero elements. The scalar field may be either real or complex, except as explicitly stated otherwise.

If T is a linear operator with domain and range in X, we adopt the following notations as a convenience in stating some of the next few theorems:

$$T_\lambda = \lambda - T, \qquad \mathscr{R}_\lambda = \text{range of } T_\lambda,$$

$M(\lambda) =$ norm of T_λ^{-1} as an operator on $\mathscr{R}_\lambda$ if T_λ^{-1} exists and is continuous. Our first fundamental results are as follows:

Theorem 5.1–A. *Suppose μ is such that $\mu - T$ has a continuous inverse. Then $\lambda - T$ has a continuous inverse if $|\lambda - \mu| M(\mu) < 1$. Moreover, $\overline{\mathscr{R}_\lambda}$ is not a proper subset of $\overline{\mathscr{R}_\mu}$.*

PROOF. Suppose $x \in \mathscr{D}(T)$. Then $T_\lambda x = \lambda x - Tx = (\lambda - \mu)x + T_\mu x$, and so

$$\|T_\lambda x\| \geqslant \|T_\mu x\| - |\lambda - \mu| \, \|x\|.$$

Now $\|x\| = \|T_\mu^{-1} T_\mu x\| \leqslant M(\mu)\|T_\mu x\|$. Therefore

$$\begin{aligned} M(\mu)\|T_\lambda x\| &\geqslant M(\mu)\|T_\mu x\| - |\lambda - \mu| M(\mu)\|x\| \\ &\geqslant \{1 - |\lambda - \mu| M(\mu)\}\|x\|. \end{aligned}$$

Because of the inequality assumed in the theorem, this last inequality shows that $x = 0$ if $T_\lambda x = 0$; hence T_λ^{-1} exists. If we write $y = T_\lambda x$, or $x = T_\lambda^{-1} y$, we see that

$$\|T_\lambda^{-1} y\| \leqslant \frac{M(\mu)\|y\|}{1 - |\lambda - \mu| M(\mu)}. \tag{5.1–1}$$

This shows that T_λ^{-1} is continuous. To prove that $\overline{\mathscr{R}_\lambda}$ is not a proper subset of $\overline{\mathscr{R}_\mu}$, assume the contrary. If θ is chosen so that $|\lambda - \mu| M(\mu) < \theta < 1$, there must exist an element $y_0 \in \overline{\mathscr{R}_\mu}$ such that $\|y_0\| = 1$ and $\|y - y_0\| \geqslant \theta$ if $y \in \overline{\mathscr{R}_\lambda}$ (by Theorem 3.12–E). Now choose $y_n \in \mathscr{R}_\mu$ so that $y_n \to y_0$. Let $x_n = T_\mu^{-1} y_n$, $y_n = T_\mu x_n$. Then $T_\lambda x_n = (\lambda - \mu)x_n + T_\mu x_n$, whence

$$\|T_\lambda x_n - T_\mu x_n\| = |\lambda - \mu| \, \|x_n\| \leqslant |\lambda - \mu| M(\mu)\|y_n\|.$$

But $T_\lambda x_n \in \mathscr{R}_\lambda$, and so

$$\theta \leqslant \|y_0 - T_\lambda x_n\| \leqslant \|y_0 - T_\mu x_n\| + \|T_\mu x_n - T_\lambda x_n\|,$$

or

$$\theta \leqslant \|y_0 - y_n\| + |\lambda - \mu| M(\mu) \|y_n\|.$$

If we now let $n \to \infty$ we obtain the contradiction $\theta \leqslant |\lambda - \mu| M(\mu)$. This finishes the proof.

Theorem 5.1–B. *The resolvent set $\rho(T)$ is open, and hence the spectrum $\sigma(T)$ is closed.*

PROOF. If $\mu \in \rho(T)$, T_μ^{-1} exists and is continuous, and $\overline{\mathscr{R}_\mu} = X$. Theorem 5.1–A shows that, if λ is sufficiently near μ, then T_λ^{-1} also exists and is continuous, and $\overline{\mathscr{R}_\lambda}$ is not a proper subset of X. Hence $\overline{\mathscr{R}_\lambda} = X$. Then $\lambda \in \rho(T)$.

It was noted in § 5.0 that, if X is complete and T is closed, the range of $\lambda - T$ is all of X when $\lambda \in \rho(T)$. For certain results about the inverse operator T_λ^{-1} we shall assume directly that $\mathscr{R}_\lambda = X$ when $\lambda \in \rho(T)$, since this is the essential condition for obtaining these results. When $\lambda \in \rho(T)$, we shall regularly denote T_λ^{-1} by R_λ; as stated in § 5.0, R_λ is called the *resolvent operator* (of T). The following theorem is of fundamental importance.

Theorem 5.1–C. *Suppose T is such that $\mathscr{R}_\lambda = X$ if $\lambda \in \rho(T)$. Then, if λ and μ are any two points in $\rho(T)$, R_λ and R_μ satisfy the relations*

$$R_\lambda - R_\mu = (\mu - \lambda) R_\lambda R_\mu, \tag{5.1–2}$$

$$R_\lambda R_\mu = R_\mu R_\lambda. \tag{5.1–3}$$

If $\mu \in \rho(T)$ and $|\mu - \lambda|\, \|R_\mu\| < 1$, then $\lambda \in \rho(T)$ and

$$R_\lambda = \sum_0^\infty (\mu - \lambda)^n R_\mu^{n+1}, \tag{5.1–4}$$

the series converging according to the metric in $[X]$. As a function on $\rho(T)$ to $[X]$, R_λ has derivatives of all orders, with

$$\frac{d^n}{d\lambda^n} R_\lambda = (-1)^n n! R_\lambda^{n+1}. \tag{5.1–5}$$

PROOF. To prove (5.1–2), suppose $y \in X$ and $x = R_\mu y$, so that $y = T_\mu x$. Then, since

$$T_\mu x - T_\lambda x = \mu x - Tx - (\lambda x - Tx) = (\mu - \lambda)x,$$

we have

$$y - T_\lambda R_\mu y = (\mu - \lambda) R_\mu y.$$

Hence, applying R_λ to both sides, we obtain

$$R_\lambda y - R_\mu y = (\mu - \lambda)R_\lambda R_\mu y.$$

This proves (5.1–2). By symmetry the result holds with λ and μ exchanged; from this we conclude that (5.1–3) holds.

We know from Theorem 5.1–A that $\mu \in \rho(T)$ and $|\lambda - \mu| \, \|R_\mu\| < 1$ imply $\lambda \in \rho(T)$. To prove (5.1–4) with this condition on λ and μ we use (5.1–2), (5.1–3), and induction to prove that

$$R_\lambda = \sum_{k=0}^{n} (\mu - \lambda)^k R_\mu^{k+1} + (\mu - \lambda)^{n+1} R_\mu^{n+1} R_\lambda$$

if $n \geqslant 0$. Thus (5.1–4) is equivalent to

$$\lim_{n\to\infty} |\mu - \lambda|^{n+1} \|R_\mu^{n+1} R_\lambda\| = 0. \tag{5.1–6}$$

Since $\|R_\mu^{n+1} R_\lambda\| \leqslant \|R_\mu\|^{n+1} \|R_\lambda\|$, and $|\mu - \lambda| \, \|R_\mu\| < 1$, we see that (5.1–6) is true.

To prove (5.1–5) when $n = 1$, perhaps the easiest method is to start with (5.1–2) and show that

$$\|(R_\lambda - R_\mu)/(\lambda - \mu) + R_\mu{}^2\| = \|(\lambda - \mu)R_\lambda R_\mu{}^2\| \leqslant |\lambda - \mu| \, \|R_\lambda\| \, \|R_\mu{}^2\|.$$

Then, from (5.1–1) we see that when $|\lambda - \mu|$ is sufficiently small

$$\|R_\lambda\| \leqslant \|R_\mu\|(1 - |\lambda - \mu| \, \|R_\mu\|)^{-1}.$$

Hence $(R_\lambda - R_\mu)/(\lambda - \mu) \to - R_\mu{}^2$ as $\lambda \to \mu$. For higher values of n, (5.1–5) may be proved by induction. See problem 2.

If $\lambda - T$ has a continuous inverse, λ need not be in $\rho(T)$, for the range of $\lambda - T$ may not be dense in X (for an example see problem 4). The following theorem tells us that, if T is closed and X is complete, such a value of λ must be at a positive distance from $\rho(T)$.

Theorem 5.1–D. *Suppose that X is complete and that T is closed. Suppose $\mu - T$ has a continuous inverse and that $\mu \in \overline{\rho(T)}$. Then $\mu \in \rho(T)$.*

PROOF. Choose $\lambda_n \in \rho(T)$ so that $\lambda_n \to \mu$. For convenience let A_n denote R_λ when $\lambda = \lambda_n$. We know that $\mathscr{R}_\lambda = X$ when $\lambda \in \rho(T)$. Hence, by (5.1–2),

$$\|A_n - A_m\| \leqslant |\lambda_n - \lambda_m| \, \|A_n\| \, \|A_m\|.$$

It follows from (5.1–1) that $\|A_n\|$ is bounded as $n \to \infty$. Hence, since

$[X]$ is complete, we see that there exists some $A \in [X]$ such that $\|A_n - A\| \to 0$. If $\lambda \in \rho(T)$ we have

$$A_n - R_\lambda = (\lambda - \lambda_n)R_\lambda A_n,$$

and so

$$A = R_\lambda + (\lambda - \mu)R_\lambda A. \tag{5.1–7}$$

This shows that the range of A lies in the range of R_λ, which is $\mathscr{D}(T)$. We shall show that $T_\mu Ay = y$ if $y \in X$, thus proving that the range of T_μ is all of X, whence $\mu \in \rho(T)$. Now, if $\lambda \in \rho(T)$,

$$T_\mu Ay = (\mu - \lambda)Ay + T_\lambda Ay. \tag{5.1–8}$$

By (5.1–7) we have $T_\lambda Ay = y + (\lambda - \mu)Ay$. When this is substituted in (5.1–8) we obtain $T_\mu Ay = y$, and the proof is finished.

PROBLEMS

1. Prove that $M(\lambda)$ is continuous on the open set of values of λ for which $\lambda - T$ has a continuous inverse. Hint: use (5.1–1) and show that, if λ and μ are in the open set referred to and $|\lambda - \mu|$ is sufficiently small, then

$$\frac{M(\mu)}{1 + |\lambda - \mu|M(\mu)} \leqslant M(\lambda) \leqslant \frac{M(\mu)}{1 - |\lambda - \mu|M(\mu)}.$$

2. To complete the proof of (5.1–5) by induction it suffices to prove that $(d/d\lambda)R_\lambda{}^n = -nR_\lambda^{n+1}$. Do this (at $\lambda = \mu$) by factoring $R_\lambda{}^n - R_\mu{}^n$.

3. Suppose that, for each λ in a nonempty set S of scalars, A_λ is a linear operator on X into X such that $A_\lambda - A_\mu = (\mu - \lambda)A_\lambda A_\mu$ if $\lambda, \mu \in S$. Suppose also that $A_\lambda{}^{-1}$ exists for at least one $\lambda \in S$. Show that $A_\lambda{}^{-1}$ exists for every $\lambda \in S$, that all the operators A_λ have the same range $\mathscr{D}$, and that there exists a linear operator T on $\mathscr{D}$ into X such that $\lambda - T = A_\lambda{}^{-1}$ for each $\lambda \in S$. Show further that, if A_λ is closed for at least one λ, then T is closed and A_λ is closed for every $\lambda \in S$.

4. Consider the operator $T \in [l^1]$ defined by $y = Tx$, where $y = \{\eta_k\}$, $x = \{\xi_k\}$, $\eta_1 = 0$, and $\eta_k = -\xi_{k-1}$ if $k \geqslant 2$. Show that $(\lambda - T)^{-1}$ exists for all λ, that $\rho(T)$ consists of all λ for which $|\lambda| > 1$, and that $\|R_\lambda\| = (|\lambda| - 1)^{-1}$. Use problem 5, § 4.51. If $|\lambda| \leqslant 1$ the range of $\lambda - T$ is not dense in l^1. The inverse is continuous if $|\lambda| < 1$, but not if $|\lambda| = 1$.

5. Let $\hat{X}$ be the completion of X (X itself if X is complete). Assume $\rho(T)$ not empty, and for each $\lambda \in \rho(T)$ let A_λ be the unique extension of $T_\lambda{}^{-1}$ to all of $\hat{X}$. Show that $A_\lambda - A_\mu = (\mu - \lambda)A_\lambda A_\mu$ if λ and μ are in $\rho(T)$. We shall say here that T has a closed linear extension if there exists a closed linear operator T_1, with domain and range in $\hat{X}$, which is an extension of T. Show that such a closed linear extension exists if and only if A_λ has an inverse for some $\lambda \in \rho(T)$ (and hence for *all* such λ, by problem 3). The next problem shows that T can fail to have a closed linear extension.

6. Take $X = l^1$. Let $\mathscr{D}(T)$ be the set of $x = \{\xi_n\}$ such that $\xi_n = 0$ except for a finite number of indices. Define $Tx = y$ by $\eta_k = k^{-2}\sum_{i=k}^{\infty} i^2\xi_i$. Then T defines a one-to-one mapping of $\mathscr{D}(T)$ onto itself; T^{-1} is continuous and T is discontinuous. The extension A of T^{-1} to all of X is defined by $Ay = x$, where $k^2\xi_k = k^2\eta_k - (k+1)^2\eta_{k+1}$. But A has no inverse. Hence T has no closed linear extension (see problem 5).

7. Let T be a linear operator with domain and range in X (not necessarily complete). Let G be the graph of T and $\bar{G}$ the closure of G (in the product space $X \times X$). If there is some closed linear operator T_0 with domain and range in X, such that T_0 is an extension of T, then $\bar{G}$ is the graph of a uniquely defined "minimal" closed extension of T, which we denote by $\bar{T}$; $\mathscr{D}(T)$ and $\mathscr{R}(T)$ are dense in $\mathscr{D}(\bar{T})$ and $\mathscr{R}(\bar{T})$, respectively. If X is complete and T is continuous on $\mathscr{D}(T)$, $\bar{G}$ is the graph of an operator $\bar{T}$ which is continuous, with $\mathscr{D}(\bar{T}) = \overline{\mathscr{D}(T)}$.

8. If X is complete and if T has a closed linear extension, the operator $\bar{T}$ (see problem 7) has the same resolvent set and spectrum as T.

5.2 The Spectrum of a Bounded Linear Operator

In all of the theorems of this section we assume that $T \in [X]$. The main results are concerned with the finding of the smallest constant r such that $|\lambda| \leqslant r$ if $\lambda \in \sigma(T)$, and with the expression of $(\lambda - T)^{-1}$ in terms of λ and T when $|\lambda| > r$.

Theorem 5.2–A. *If $T \in [X]$ and $|\lambda| > \|T\|$, $(\lambda - T)^{-1}$ exists and is continuous, and*

$$(\lambda - T)^{-1}y = \sum_{1}^{\infty} \lambda^{-n}T^{n-1}y \tag{5.2–1}$$

for each y in the range of $\lambda - T$. If X is complete, $|\lambda| > \|T\|$ implies $\lambda \in \rho(T)$, and then

$$(\lambda - T)^{-1} = R_\lambda = \sum_{1}^{\infty} \lambda^{-n}T^{n-1}, \tag{5.2–2}$$

the series converging in the space $[X]$.

PROOF. If $|\lambda| > \|T\|$, we have

$$\|\lambda x - Tx\| \geqslant |\lambda|\,\|x\| - \|Tx\| \geqslant (|\lambda| - \|T\|)\,\|x\|.$$

The existence and continuity of $(\lambda - T)^{-1}$ now follows by Theorem 3.1–B. If $y = (\lambda - T)x$, we find by induction that

$$x = \lambda^{-1}y + \cdots + \lambda^{-n}T^{n-1}y + \lambda^{-n}T^n x.$$

Formula (5.2–1) follows from this, because $\lambda^{-n}T^n x \to 0$ when $|\lambda| > \|T\|$. The assertion about (5.2–2), when X is complete, follows by taking $A = \lambda^{-1}T$ in Theorem 4.1–C.

We see from Theorem 5.2–A that, when X is complete, $\lambda \in \sigma(T)$ implies $|\lambda| \leqslant \|T\|$. Since $\sigma(T)$ is closed, this shows that $\sigma(T)$ is compact when $T \in [X]$ and X is complete. If we are given some compact set, there is some bounded linear operator having this compact set as its spectrum. See problem 1.

Before going further we note this useful fact: *If* $T \in [X]$ *and if the series* $\sum_{1}^{\infty} \lambda^{-n}T^{n-1}$ *converges in* $[X]$ *for some value of* λ, *then* $\lambda \in \rho(T)$ *and the operator defined by the series is* R_λ. *For, denoting the series by* A, *it is easily seen that* $(\lambda - T)A = A(\lambda - T) = I$.

For the remainder of this section we need to assume that X is a complex Banach space. The reasons for this are not superficial. We need the theory of analytic functions of a complex variable, with values in $[X]$. We cannot relate the convergence of the series (5.2–2) to the extent of the set $\rho(T)$ unless we think of λ as a complex variable, for somewhat the same reason that the radius of convergence of the power series expansion of a real analytic function cannot be discovered merely by looking for singularities of the function on the real axis.

In §§ 4.4 and 4.42 we commented on the concept of an analytic function of a complex variable with values in a complex Banach space. We must now make some use of the theory of such functions. A great deal of the standard classical theory can be taken over intact, proofs and all. Cauchy's integral theorem, the integral formulas for a function and its derivatives, Taylor's theorem, Laurent's theorem, Liouville's theorem, and many other theorems retain their validity. The proofs are just as in classical theory, except that norms replace absolute values. Some further discussion of these matters is indicated in the problems. For our immediate purposes we need the theorems of Liouville and Laurent.

The key fact is that, when T is closed and X is a complex Banach space, R_λ depends analytically on λ as λ varies in $\rho(T)$. This is true by Theorem 5.1–C.

Theorem 5.2–B. *If* $T \in [X]$ *and* X *is a complex Banach space,* $\sigma(T)$ *is not empty.*

PROOF. It appears from the proof of Theorem 5.2–A that $\|R_\lambda\| \leqslant (|\lambda| - \|T\|)^{-1}$ if $|\lambda| > \|T\|$. Hence $\|R_\lambda\| \to 0$ as $|\lambda| \to \infty$. If $\sigma(T)$ were empty, it would follow that R_λ is analytic and bounded on the whole plane. But then it would be constant, by Liouville's theorem, and the constant would be the zero operator. This is impossible, by the fact that R_λ sets up a one-to-one mapping of X onto itself, for X is assumed to have some nonzero elements (see the beginning of § 5.1).

Theorem 5.2–B remains valid even if X is not complete, as may be shown by an argument involving the completion $\hat{X}$ and an extension of the operator. See problem 2.

Definition. Supposing $\sigma(T)$ nonempty and bounded, we define

$$r_\sigma(T) = \sup_{\lambda\in\sigma(T)} |\lambda|, \tag{5.2–3}$$

and call $r_\sigma(T)$ the *spectral radius* of T.

Theorem 5.2–C. *If $T \in [X]$ and X is a complex Banach space, the resolvent is given by*

$$R_\lambda = \sum_1^\infty \lambda^{-n} T^{n-1} \tag{5.2–4}$$

if $|\lambda| > r_\sigma(T)$. This series also represents R_λ if the series converges and $|\lambda| = r_\sigma(T)$. The series diverges if $|\lambda| < r_\sigma(T)$.

PROOF. We know that R_λ is analytic when $|\lambda| > r_\sigma(T)$. Hence it has a unique Laurent expansion in positive and negative powers of λ, convergent when $|\lambda| > r_\sigma(T)$. Now, we already know that (5.2–4) is valid when $|\lambda| > \|T\|$. By the uniqueness, then, this must be the Laurent expansion (it may also be called the Taylor expansion about the point $\lambda = \infty$). The second assertion in the theorem follows from the italicized statement in the second paragraph after the proof of Theorem 5.2–A. For the same reason, the series (5.2–4) cannot converge if $\lambda \in \sigma(T)$. Hence it cannot converge at λ_0 if $|\lambda_0| < r_\sigma(T)$, because, if it did, it would follow, as in the general theory of power series, that the series converges when $|\lambda| > |\lambda_0|$, and so, in particular, for some $\lambda \in \sigma(T)$.

Theorem 5.2–C enables us to write a formula for the spectral radius of T. If we consider the series (5.2–4) as a power series in λ^{-1}, the standard formula for the radius of convergence of a power series tells us that

$$r_\sigma(T) = \limsup_{n\to\infty} \|T^n\|^{1/n}. \tag{5.2–5}$$

Actually, as we shall presently prove, $\{\|T^n\|^{1/n}\}$ is a convergent sequence, so that the limit superior in (5.2–5) is a limit. To prove this, we first prove what is called the *spectral-mapping theorem* for polynomials. Suppose

$$F(\lambda) = \alpha_n\lambda^n + \alpha_{n-1}\lambda^{n-1} + \cdots + \alpha_0$$

is a polynomial with complex coefficients. If $T \in [X]$, positive integral powers of T have a clear meaning, and we define

$$F(T) = \alpha_n T^n + \alpha_{n-1} T^{n-1} + \cdots + \alpha_0 I.$$

By the rules of algebra for operators it is clear that, if the polynomial $F(\lambda)$ is factored, there is a corresponding factored form of $F(T)$.

Now $F(T) \in [X]$, and we can consider its spectrum.

Theorem 5.2–D. *Suppose $T \in [X]$, where X is a complex Banach space. If F is a polynomial, the spectrum of $F(T)$ consists precisely of those points μ such that $F(\lambda) = \mu$ for some $\lambda \in \sigma(T)$. In symbolic form, $\sigma\{F(T)\} = F\{\sigma(T)\}$.*

PROOF. We can assume that $n \geqslant 1$ and $\alpha_n = 1$, leaving the case $n = 0$ to the reader. For a fixed μ let the zeros of $F(\lambda) - \mu$ be $\beta_1, \cdots, \beta_n$, so that

$$F(T) - \mu I = (T - \beta_1) \cdots (T - \beta_n). \tag{5.2–6}$$

If $T - \beta_1, \cdots, T - \beta_n$ each have continuous inverses defined on all of X, so does $F(T) - \mu I$, the inverse of the latter being the product of the inverses of the former in the reverse order. Hence, if $\mu \in \sigma\{F(T)\}$, there must be some β_k such that $\beta_k \in \sigma(T)$. Since $F(\beta_k) = \mu$, this shows that $\sigma\{F(T)\} \subset F\{\sigma(T)\}$. Suppose, on the other hand, that some β_k, say β_1, is in $\sigma(T)$. If $T - \beta_1$ has an inverse, the range of $T - \beta_1$ is not all of X, and (5.2–6) shows that the range of $F(T) - \mu I$ is likewise not all of X; hence $\mu \in \sigma\{F(T)\}$. If $T - \beta_1$ has no inverse, we see by exchanging the positions of the factors $T - \beta_1$ and $T - \beta_n$ in (5.2–6) that $F(T) - \mu I$ also has no inverse, and again $\mu \in \sigma\{F(T)\}$. This argument works just as well for any β_k as for β_1, and so the proof is complete.

Theorem 5.2–E. *Suppose $T \in [X]$, where X is a complex Banach space. Then $r_\sigma(T) \leqslant \|T^n\|^{1/n}$ for every positive integer n. Also, $\|T^n\|^{1/n}$ converges to $r_\sigma(T)$ as $n \to \infty$.*

PROOF. Theorem 5.2–D shows that $\sigma(T^n)$ consists of the nth powers of points of $\sigma(T)$. Hence $r_\sigma(T^n) = [r_\sigma(T)]^n$. We know (Theorem 5.2–A) that $r_\sigma(T^n) \leqslant \|T^n\|$. Hence $r_\sigma(T) \leqslant \|T^n\|^{1/n}$. It follows that

$$r_\sigma(T) \leqslant \liminf_{n\to\infty} \|T^n\|^{1/n}.$$

This, together with (5.2–5), leads to the final conclusion of the theorem.

PROBLEMS

1. Let S be a compact set in the plane. Let $\{\alpha_n\}$ be a sequence of points of S everywhere dense in S. Define $T \in [l^2]$ by $Tx = y$, where $x = \{\xi_n\}$,

$y = \{\alpha_n \xi_n\}$. Then $S = \sigma(T)$. Each α_n is an eigenvalue; for the other points $\lambda \in S$, the range of $\lambda - T$ is dense in l^2 and the inverse is discontinuous.

2. To prove Theorem 5.2–B when X is incomplete, let $\hat{T}$ be the unique linear extension of T to all of the completion $\hat{X}$. Prove that $\rho(T) \subset \rho(\hat{T})$, and hence $\sigma(\hat{T}) \subset \sigma(T)$, in two steps: *(a)* if $\lambda \in \rho(T)$, $\lambda - \hat{T}$ has an inverse; *(b)* the range of $\lambda - \hat{T}$ is all of $\hat{X}$, and hence the inverse is continuous.

3. Instead of proving theorems about Banach-space-valued analytic functions by repeating the classical proofs, it is sometimes possible to make use of linear functionals. For instance, to prove Liouville's theorem, assume that f, with values in X, is differentiable at all points of the complex plane and bounded. Then $x'[f(\lambda)]$ is a scalar-valued function with the same properties, for each $x' \in X'$ and is therefore constant. Thus $x'[f(\lambda) - f(0)] = 0$ for each x' and each λ. But then $f(\lambda) \equiv f(0)$. Cauchy's theorem may be proved in a similar way, by noting that $x'\left\{\int_C f(\lambda)\, d\lambda\right\} = \int_C x'[f(\lambda)]\, d\lambda$.

4. If $X = C[a, b]$ and T is a Volterra-type integral operator (see § 4.11), $\sigma(T)$ is the single point $\lambda = 0$.

5. If T does not belong to $[X]$, $\sigma(T)$ may not be compact. Example 3, § 1.5 illustrates a situation in which the spectrum of a certain differential operator consists of the points $-n^2$, $n = 0, 1, 2, \cdots$. Example 4, § 1.5 shows a differential operator whose spectrum is composed of all λ whose real part is $\leqslant 0$.

6. Suppose $A, B, C \in [X]$, $\lambda \neq 0$, and $C(\lambda - BA) = (\lambda - BA)C = I$. Let $D = \lambda^{-1} + \lambda^{-1}ACB$. Then $D(\lambda - AB) = (\lambda - AB)D = I$. Thus, when X is complete, the nonzero points of $\rho(AB)$ and $\rho(BA)$ are the same. Hence AB and BA have the same spectral radius.

7. If X is a complex Banach space and $AB = BA$, where A and B are in $[X]$, then $r_\sigma(AB) \leqslant r_\sigma(A) r_\sigma(B)$.

8. Suppose $A \in [X]$, $A^n \neq 0$ for all n, and let $\|A^n\|^{1/n}$ be monotonic in n. Show that $\|A^{n+1}\|/\|A^n\| \leqslant \|A^n\|^{1/n}$, whence it can be inferred that $\overline{\lim}\, (\|A^{n+1}\|/\|A^n\|) = \lim \|A^n\|^{1/n}$. For a case in which $\|A^n\|^{1/n}$ is not monotone, let $A \in [l^1]$ be defined by $Ax = y$, where $\eta_1 = 0$, $\eta_{2k} = \xi_{2k-1}$, $\eta_{2k+1} = 2\xi_{2k}$, $k = 1, 2, \cdots$.

5.3 Subdivisions of the Spectrum

We can classify the various values of the parameter λ according to the state of the operator $\lambda - T$, using the definitions of "states" as made in § 4.71. We say that λ is in one of the classes I_1, I_2, $\cdots$, III_2, III_3 if $\lambda - T$ is in the corresponding state, as an operator on $\mathscr{D}(T)$ into X. For this classification we do not insist that T be continuous or even closed. According to the definitions, $\lambda \in \rho(T)$ if and only if λ is in class I_1 or II_1. It has been customary to group the remaining classes as follows, thus dividing the spectrum into three mutually exclusive parts:

Classes I_2 and II_2 = the continuous spectrum, denoted by $C\sigma(T)$;
Classes III_1 and III_2 = the residual spectrum, denoted by $R\sigma(T)$;

Classes I_3, II_3, and III_3 = the point spectrum (eigenvalues), denoted by $P\sigma(T)$.

When $T \in [X]$, the operator T' is defined, and $(\lambda - T)' = \lambda - T'$. The results of § 4.71, and the state diagram in particular, can be applied to show exactly how the classification of λ in relation to T affects its classification in relation to T', and vice versa. We note the following result:

Theorem 5.3–A. *If $T \in [X]$, then T and T' have the same resolvent set and the same spectrum.*

For illustrative purposes we shall discuss two operators, each of which may be considered as a bounded operator on l^p into l^p for any selected value of p, $1 \leqslant p \leqslant \infty$. We shall analyze the spectrum of each operator and see how the resulting classification of each spectral value λ depends on the value of p. Some of the details are left for the problems. We always write $x = \{\xi_n\}$, $y = \{\eta_n\}$, $n = 1, 2, \cdots$.

Example 1. Let T on l^p into l^p be defined by the infinite matrix

$$\left\|\begin{matrix} 0 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 0 & \\ 0 & 0 & 0 & 1 & 0 & \\ & & & \cdot & & \\ & & & & \cdot & \\ \cdots & & & & \cdot & \end{matrix}\right\|$$

so that $(\lambda - T)x = y$ means $\eta_k = \lambda\xi_k - \xi_{k+1}$, $k = 1, 2, \cdots$. It is easy to see that $\|T\| = 1$ for each value of p. Hence $|\lambda| > 1$ implies $\lambda \in \rho(T)$. It is readily seen that when $p = \infty$ and $|\lambda| \leqslant 1$, λ is an eigenvalue; the corresponding eigenmanifold is generated by the vector $(1, \lambda, \lambda^2, \cdots)$. When $1 \leqslant p < \infty$ we have this same eigenvalue and eigenmanifold if $|\lambda| < 1$. But $(\lambda - T)^{-1}$ exists if $|\lambda| = 1$ and $1 \leqslant p < \infty$, for $(1, \lambda, \lambda^2, \cdots)$ is not an element of l^p in this case. Since $\sigma(T)$ is closed, we see that, for each value of p, $\sigma(T)$ is the set $\{\lambda : |\lambda| \leqslant 1\}$. When $|\lambda| = 1$ and $1 \leqslant p < \infty$, the inverse cannot be continuous, by Theorem 5.1–D.

To investigate the range of $\lambda - T$ we find by induction from $\eta_k = \lambda\xi_k - \xi_{k+1}$ that

$$\xi_{k+1} = \lambda^k\xi_1 - \lambda^{k-1}\eta_1 - \lambda^{k-2}\eta_2 - \cdots - \eta_k$$

and that

$$\xi_1 = \lambda^{-1}\eta_1 + \cdots + \lambda^{-k}\eta_k + \lambda^{-k}\xi_{k+1} \text{ if } \lambda \neq 0.$$

Note that $\lambda^{-k}\xi_{k+1} \to 0$ as $k \to \infty$ if $|\lambda| \geqslant 1$ and $1 \leqslant p < \infty$ or if $|\lambda| > 1$ and $p = \infty$. This enables us to find x in terms of y when $y = (\lambda - T)x$.

In particular, we get the formulas for the resolvent operator when $|\lambda| > 1$. When $|\lambda| = 1$ and $1 \leqslant p < \infty$, the range of $\lambda - T$ is dense in l^p, for it is easily seen to contain y if the number of nonzero components of y is finite. The range cannot be all of l^p in this case, however. Why not? We leave it until later to show that the range is *not* dense in l^p for the case $|\lambda| = 1$ and $p = \infty$. When $\lambda = 0$, the range of $\lambda - T$ is obviously all of l^p, for each p. When $0 < |\lambda| < 1$ and y is in the range of $\lambda - T$, any x such that $(\lambda - T)x = y$ is given by $x = \xi_1(1, \lambda, \lambda^2, \cdots) - \lambda^{-1}z$, where $z = \{\zeta_k\}$, $\zeta_1 = 0$, and

$$\zeta_{k+1} = \lambda^k\eta_1 + \lambda^{k-1}\eta_2 + \cdots + \lambda\eta_k \qquad \text{if } k \geqslant 1. \tag{5.3–1}$$

After discussing Example 2 we shall see that the range of $\lambda - T$ is all of l^p in this case, for each p.

Example 2. Let A on l^p into l^p be defined by the infinite matrix

$$\begin{Vmatrix} 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & \\ 0 & 0 & 1 & 0 & \\ & & & \ddots & \\ \cdots & & & & \ddots \end{Vmatrix}$$

This is the transpose of the matrix of Example 1. When $p = \infty$, A can be identified with the conjugate operator T', T being the operator of Example 1, with $p = 1$. If $1 < p < \infty$ and $p' = p/(p - 1)$, A for p' can be identified with the conjugate T' of T for p. Finally, T for $p = \infty$ can be identified with the conjugate A' of A for $p = 1$. These remarks, and the state diagram of § 4.71, will help us in analyzing the spectrum of A. In particular, the relations between an operator and its conjugate assure us that $\|A\| = 1$ (which we can of course see directly) and that $\sigma(A) = \sigma(T) = \{\lambda : |\lambda| \leqslant 1\}$.

The equation $(\lambda - A)x = y$ is expressed by the equations $\eta_1 = \lambda\xi_1$, $\eta_{k+1} = -\xi_k + \lambda\xi_{k+1}$, $k \geqslant 1$. When $\lambda = 0$ we see that the inverse of $\lambda - A$ exists and is continuous, for $\|Ax\| = \|x\|$. The range is not dense in l^p, however, for $y = -Ax$ implies $\eta_1 = 0$; the range is a proper closed subspace of l^p. When $\lambda \neq 0$ the inverse of $\lambda - A$ exists; $x = (\lambda - A)^{-1}y$, for y in the range of $\lambda - A$, is expressed by

$$\xi_k = \lambda^{-k}\eta_1 + \lambda^{1-k}\eta_2 + \cdots + \lambda^{-1}\eta_k. \tag{5.3–2}$$

In particular, these equations define the resolvent of A when $|\lambda| > 1$.

We can use equations (5.3–2) in discussing the range of $\lambda - T$. A scrutiny of (5.3–1) and (5.3–2) shows that the z of (5.3–1) can be written in the form $z = A(\lambda^{-1} - A)^{-1}y$. The solutions of $(\lambda - T)x = y$ are then $x = \xi_1(1, \lambda, \lambda^2, \cdots) - \lambda^{-1}A(\lambda^{-1} - A)^{-1}y$, where ξ_1 is arbitrary. If $0 < |\lambda| < 1$, this formula is applicable for every y in l^p, for all p, and so we see that the range of $\lambda - T$ is all of l^p in this case. The result is valid for $0 \leqslant |\lambda| < 1$ if we write it in the form $x = \xi_1(1, \lambda, \lambda^2, \cdots) + A(\lambda A - I)^{-1}y$.

Next we show that, for $p = \infty$, the range of $\lambda - A$ is not dense in l^∞ if $|\lambda| = 1$. In fact, we show that y is not in the range if $\|y - w\| = \epsilon < 1$, where $w = (\lambda^{-1}, \lambda^{-2}, \lambda^{-3}, \cdots)$. If we assume that $x \in l^\infty$ and $(\lambda - A)x = y$, we can write $y = w + (\theta_1, \theta_2, \cdots)$, where $|\theta_n| \leqslant \epsilon$, and we have, from (5.3–2),

$$\xi_n = n\lambda^{-n-1} + \lambda^{-n-1}(\lambda\theta_1 + \lambda^2\theta_2 + \cdots + \lambda^n\theta_n),$$

so that

$$|n^{-1}\xi_n - \lambda^{-n-1}| = n^{-1}|\lambda\theta_1 + \cdots + \lambda^n\theta_n| \leqslant \epsilon.$$

Letting $n \to \infty$, we obtain the contradiction $1 \leqslant \epsilon$.

An argument of a similar kind will show for the case $p = \infty$, $|\lambda| = 1$, that the range of $\lambda - T$ is not dense in l^∞. If $y = (\lambda - T)x$, it may be verified that $y = (\lambda\xi_1, 0, 0, \cdots) - (\lambda^{-1} - A)T(\lambda x)$, and from this it may be shown that y is not in the range of $\lambda - T$ if $\|y - w\| = \epsilon < 1$, where $w = (\lambda, \lambda^2, \lambda^3, \cdots)$.

It is now possible to make a complete classification of the points of the spectra of T and A for all values of p. See the problems.

PROBLEMS

1. For the T of Example 1 there is the following classification of points of $\sigma(T)$. (*a*) $\lambda = 0{:}I_3$; (*b*) $0 < |\lambda| < 1{:}I_3$; (*c*) $|\lambda| = 1{:}\mathrm{II}_2$ if $1 \leqslant p < \infty$, and III_3 if $p = \infty$.

2. For the A of Example 2 there is the following classification of points of $\sigma(A)$. (*a*) $|\lambda| < 1{:}\mathrm{III}_1$; (*b*) $|\lambda| = 1{:}\mathrm{II}_2$ if $1 < p < \infty$, and III_2 if $p = 1$ or ∞. Some of these results are obtained by using problem 1 and the state diagram.

3. When $p = 1$ and $0 \leqslant |\lambda| < 1$ in Example 2, show that $\|(\lambda - A)x\| \geqslant (1 - |\lambda|)\|x\|$, so that $(\lambda - A)^{-1}$ is continuous.

4. If $(\lambda - A)x = y$ in Example 2, show that $\eta_1 + \lambda\eta_2 + \cdots + \lambda^k\eta_{k+1} = \lambda^{k+1}\xi_{k+1}$. This may be used to show that the range of $\lambda - A$ is not dense in l^p if $|\lambda| < 1$ and $p = \infty$, or if $|\lambda| \leqslant 1$ and $p = 1$.

5. If $|\lambda| = 1$ and $1 \leqslant p < \infty$ in Example 2, take $x_n = (1, \lambda^{-1}, \cdots, \lambda^{1-n}, 0, 0, \cdots)$, and show that $(\lambda - A)^{-1}$ is not continuous.

6. Let $\alpha_1, \alpha_2, \cdots$ be scalars such that $\sup_k |\alpha_k| < \infty$. Let $\beta_2, \beta_3 \cdots$ be scalars such that $\sum_2^\infty |\beta_k| < \infty$. Define $T \in [l^1]$ by $\eta_1 = \sum_1^\infty \alpha_i \xi_i$, $\eta_k = \beta_k \xi_1$ if $k \geqslant 2$. Discuss $\sigma(T)$ and find the resolvent operator.

7. Let $T \in [l^1]$ be defined by the infinite matrix

$$\begin{Vmatrix} 0 & 1 & 1 & 1 & \cdots \\ 1 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ & & & \ddots & \\ \cdots & & & & \end{Vmatrix}.$$

Show that $\sigma(T)$ consists of $\lambda = (1 + \sqrt{5})/2$ and all λ such that $|\lambda| \leqslant 1$. Classify the points of $\sigma(T)$. This operator has an interesting connection with the Fibonacci numbers, as may be seen by computing the matrix representation of T^n. The operator has been studied by Halberg, **1**, pages 21–29.

8. The operator $A \in [l^1]$ defined by the matrix (α_{ij}) with $\alpha_{ij} = 1$ if $|i - j| = 1$, $\alpha_{ij} = 0$ otherwise, is interesting. Its spectrum consists of the interval $-2 \leqslant \lambda \leqslant 2$ of the real axis. The points $\lambda = \pm 2$ are classified II_2; the rest of $\sigma(A)$ is classified III_2 (see Halberg, **1**, pages 29–36). We may also consider A as an element of $[l^2]$. The spectrum is the same as before, but now all points are classified II_2 (Hellinger, **1**, pages 231–232).

5.4 Reducibility

Let T be a linear operator with domain and range in the linear space X. For convenience we shall sometimes write $\mathscr{D}$ instead of $\mathscr{D}(T)$ for the domain of T. A subspace M of X is said to be *invariant* under T if $T(\mathscr{D} \cap M) \subset M$. We can then talk about the *restriction* of T to M, with M in place of X and $\mathscr{D} \cap M$ in place of $\mathscr{D}$.

Definition. The operator T is said to be *completely reduced* by the pair of subspaces (M_1, M_2) if these subspaces are linearly independent (see § 4.8) and invariant under T, such that $X = M_1 \oplus M_2$ and such that $P_i\mathscr{D} \subset \mathscr{D}$ $(i = 1, 2)$, where P_1 and P_2 are the projections defined by $P_i x = x_i$, where $x = x_1 + x_2$ $(x_i \in M_i)$ is the direct sum representation of an arbitrary x.

This definition can be extended in an obvious manner to give meaning to the statement "T is completely reduced by the set of subspaces $M_1, \cdots, M_n$."

Let us see the meaning of the foregoing concepts in terms of the matrix representation of T when X is n-dimensional and $\mathscr{D} = X$. Suppose that

M_1 is r-dimensional ($1 \leqslant r < n$), and let us choose a basis for X such that the first r elements of the basis form a basis for M_1. Then, if M_1 is invariant under T, the matrix which represents T will have all zero elements in the intersection of the first r columns and the last $n - r$ rows. If M_2 is the subspace generated by the last $n - r$ basis elements and if T is completely reduced by (M_1, M_2), all of the elements in the intersection of the first r rows and the last $n - r$ columns will also be zeros.

If an operator is completely reduced by a pair of subspaces, the operator may be studied by studying the restrictions of it to these subspaces. If these restrictions can also be completely reduced, the study of the operator may be simplified further. Evidently, if X is n-dimensional and $\mathscr{D} = X$, the greatest simplification will occur if we can find a set of n subspaces $M_1, \cdots, M_n$, each of them one dimensional, which completely reduce the operator. This is the same as asking that X have a basis $u_1, \cdots, u_n$, each element of which is an eigenvector. The matrix representing the operator will then have eigenvalues down its main diagonal, and all other elements will be zero. This situation can occur only for operators of a very special sort.

We are speaking here of finite sets of subspaces which completely reduce an operator. There are important generalizations of this concept for infinite sets of subspaces. For the present we do not try to develop such generalizations, but we mention the following concrete example.

Let $X = L^2(0, 2\pi)$. Let $\mathscr{D}$ be the subspace of X determined by those functions $x(t)$ in $\mathscr{L}^2(0, 2\pi)$ which are absolutely continuous on $[0, 2\pi]$ and are such that $x(0) = x(2\pi)$ and such that the derivative $x'(t)$ is also in $\mathscr{L}^2(0, 2\pi)$. We then define $Tx = y$, where $y(t) = x'(t)$. Let $u_n(t) = (1/\sqrt{2\pi})e^{int}$, $n = 0, \pm 1, \pm 2, \cdots$. The u_n's form a complete orthonormal set in X. Observe that $u_n \in \mathscr{D}$ and that the closed linear manifold generated by each u_n is invariant under T; the same is true of the closed linear manifold generated by any finite or infinite subset of the u_n's. Hence T is completely reduced by $(M, M^{\perp})$, where M is the closed linear manifold generated by any set of the u_n's In a certain sense T is completely reduced by the infinite set of subspaces $\{M_n\}$, where M_n is generated by u_n.

If $X = M_1 \oplus M_2$ and if P_1, P_2 are the projections determined by this direct sum representation of X, an operator T with domain $\mathscr{D}$ is completely reduced by (M_1, M_2) if and only if $P_1\mathscr{D} \subset \mathscr{D}$ and $P_1Tx = TP_1x$ when $x \in \mathscr{D}$. The corresponding conditions will then be satisfied by P_2. The simple verification of these assertions is left to the reader. It is then easily proved that, if T is completely reduced by (M_1, M_2) and if T^{-1} exists, then T^{-1} also is completely reduced by (M_1, M_2).

For formal reference we now state several theorems.

Theorem 5.4–A. *Let T be completely reduced by (M_1, M_2). Let T_k be the restriction of T to M_k. Then (a) $\mathscr{D}(T) = \mathscr{D}(T_1) \oplus \mathscr{D}(T_2)$; (b) $\mathscr{R}(T) = \mathscr{R}(T_1) \oplus \mathscr{R}(T_2)$; (c) T^{-1} exists if and only if T_1^{-1} and T_2^{-1} exist; (d) $\mathscr{R}(T) = X$ if and only if $\mathscr{R}(T_k) = M_k$ for $k = 1$ and 2.*

The proof is left to the reader.

Theorem 5.4–B. *Consider the situation of Theorem 5.4–A, supposing now that X is a normed linear space and that the projections P_1, P_2 determined by the direct sum $X = M_1 \oplus M_2$ are continuous. Then $\mathscr{R}(T)$ is dense in X if and only if $\mathscr{R}(T_1)$ and $\mathscr{R}(T_2)$ are dense in M_1 and M_2 respectively. If T^{-1} exists, it is continuous if and only if T_1^{-1} and T_2^{-1} are continuous.*

The proof is left to the reader. We remark that P_1 and P_2 will certainly be continuous if X is complete and M_1 and M_2 are closed; see problem 2, § 4.8.

When the conditions of Theorems 5.4–A and 5.4–B are satisfied, we can rephrase most of the results of these theorems as follows, using the terminology of § 4.71 regarding the classifications I, II, III and 1, 2, 3:

T is classified as 1 if and only if T_1 and T_2 are both classified as 1.
T is classified as I if and only if T_1 and T_2 are both classified as I.
T is classified as 3 if and only if at least one of T_1, T_2 is classified as 3.
T is classified as III if and only if at least one of T_1, T_2 is classified as III.

From these statements, as applied to $\lambda - T$, which is completely reduced by (M_1, M_2) if T is, we have the following theorem, verification of which is left to the reader.

Theorem 5.4–C. *Under the conditions of Theorem 5.4–B we have*

a. $\sigma(T) = \sigma(T_1) \cup \sigma(T_2)$.
b. $P\sigma(T) = P\sigma(T_1) \cup P\sigma(T_2)$.

If it is furthermore assumed that $\sigma(T_1)$ and $\sigma(T_2)$ have no points in common, it follows that

c. $C\sigma(T) = C\sigma(T_1) \cup C\sigma(T_2)$
and
d. $R\sigma(T) = R\sigma(T_1) \cup R\sigma(T_2)$.

PROBLEMS

1. Verify the assertions made in the paragraph immediately before Theorem 5.4–A.

2. Write out the proofs of Theorems 5.4–A and 5.4–B.

3. Write out the proof of Theorem 5.4–C.

4. If $T \in [X]$, where X is a normed linear space, if T is completely reduced by (M_1, M_2), and if the corresponding projections P_1, P_2 are continuous, then T' is completely reduced by $(M_1{}^0, M_2{}^0)$. See problem 4, § 4.8.

5.41 The Ascent and Descent of an Operator

In this section X denotes a linear space, and T is a linear operator with $\mathscr{D}(T)$ and $\mathscr{R}(T)$ in X. The considerations are all algebraic. We define T^n by induction, with $T^0 = I$, $T^1 = T$. If $n \geqslant 1$, $\mathscr{D}(T^n)$ is the set of x such that $x, Tx, \cdots, T^{n-1}x$ are all in $\mathscr{D}(T)$. If $n \geqslant 2$, $\mathscr{D}(T^n) = \{x : x$ and Tx are in $\mathscr{D}(T^{n-1})\}$. For convenience we write $\mathscr{D}(T^n) = \mathscr{D}_n(T)$; observe that $\mathscr{D}_0(T) = X$. In general we have $\mathscr{D}_n(T) \subset \mathscr{D}_{n-1}(T)$; the inclusion may be proper.

We generalize a previous notation and terminology by defining the *null manifold* of T^n as the set of $x \in \mathscr{D}_n(T)$ such that $T^n x = 0$. We denote this subspace of X by $\mathscr{N}(T^n)$. Evidently $\mathscr{N}(T^0) = (0)$.

Theorem 5.41–A. $\mathscr{N}(T^n) \subset \mathscr{N}(T^{n+1})$, $n = 0, 1, \cdots$. *If* $\mathscr{N}(T^k) = \mathscr{N}(T^{k+1})$ *for some* k, *then* $\mathscr{N}(T^k) = \mathscr{N}(T^n)$ *when* $n \geqslant k$.

PROOF. The first statement is evident. Suppose that $\mathscr{N}(T^n) = \mathscr{N}(T^{n+1})$ for some n and that $x \in \mathscr{N}(T^{n+2})$. Then $x \in \mathscr{D}_{n+2}(T)$ and $T^{n+2}x = 0$. Hence $Tx \in \mathscr{D}_{n+1}(T)$ and $T^{n+1}(Tx) = 0$, so that $Tx \in \mathscr{N}(T^{n+1}) = \mathscr{N}(T^n)$. Then $T^n(Tx) = T^{n+1}x = 0$. This shows that $\mathscr{N}(T^{n+2}) \subset \mathscr{N}(T^{n+1})$. The reverse inclusion also holds, and so $\mathscr{N}(T^{n+1}) = \mathscr{N}(T^{n+2})$. The second assertion of the theorem now follows by induction.

Definition. If there is some integer $n \geqslant 0$ such that $\mathscr{N}(T^n) = \mathscr{N}(T^{n+1})$, the smallest such integer is called the *ascent* of T and denoted by $\alpha(T)$. If no such integer exists we say that $\alpha(T) = \infty$.

Next we consider the ranges $\mathscr{R}(T^n)$. Note that $\mathscr{R}(T^0) = X$.

Theorem 5.41–B. $\mathscr{R}(T^{n+1}) \subset \mathscr{R}(T^n)$, $n = 0, 1, \cdots$. *If* $\mathscr{R}(T^{k+1}) = \mathscr{R}(T^k)$ *for some* k, *then* $\mathscr{R}(T^n) = \mathscr{R}(T^k)$ *when* $n \geqslant k$.

PROOF. If $y \in \mathscr{R}(T^{n+1})$, $y = T^{n+1}x$, where $x \in \mathscr{D}_{n+1}(T)$. Then $Tx \in \mathscr{D}_n(T)$ and $y = T^n(Tx)$, so that $y \in \mathscr{R}(T^n)$. This proves the first statement. Before proving the next assertion we observe that

$$(5.41\text{–}1) \qquad \mathscr{R}(T^{n+1}) = T\{\mathscr{R}(T^n) \cap \mathscr{D}(T)\}, \qquad n = 0, 1, 2, \cdots.$$

Now suppose that $\mathscr{R}(T^{n+1}) = \mathscr{R}(T^n)$ for some n, and suppose $y \in \mathscr{R}(T^{n+1})$. By (5.41–1) we can write $y = Tx$, where $x \in \mathscr{R}(T^n) \cap \mathscr{D}(T) = \mathscr{R}(T^{n+1}) \cap \mathscr{D}(T)$. Thus, again by (5.41–1), $y \in \mathscr{R}(T^{n+2})$. This shows that $\mathscr{R}(T^{n+1}) \subset \mathscr{R}(T^{n+2})$, and the proof is finished just as in the case of the previous theorem.

Definition. If there is some integer $n \geqslant 0$ such that $\mathscr{R}(T^{n+1}) = \mathscr{R}(T^n)$, the smallest such integer is called the *descent* of T and denoted by $\delta(T)$. If there is no such integer, we say that $\delta(T) = \infty$.

Observe that $\alpha(T) = 0$ if and only if T^{-1} exists and that $\delta(T) = 0$ if and only if $\mathscr{R}(T) = X$.

There are certain relations between the ascent and descent of an operator, provided that certain auxiliary conditions are satisfied.

Theorem 5.41–C. *If $\alpha(T)$ is finite and $\delta(T) = 0$, then $\alpha(T) = 0$ also.*

PROOF. Suppose that $0 < \alpha(T)$ and $\delta(T) = 0$. Then $\mathscr{R}(T) = X$ and T^{-1} does not exist. Choose $x_1 \neq 0$ so that $Tx_1 = 0$. By induction we define $x_2, x_3, \cdots$ so that $x_{n+1} \in \mathscr{D}(T)$ and $Tx_{n+1} = x_n$. Then we see that $x_{n+1} \in \mathscr{D}_{n+1}(T)$, $T^n x_{n+1} = x_1$, $T^{n+1}x_{n+1} = 0$. Thus $x_{n+1} \in \mathscr{N}(T^{n+1}) - \mathscr{N}(T^n)$. Hence $\alpha(T) = \infty$. This proves the theorem.

Theorem 5.41–D. *If $\alpha(T)$ and $\delta(T)$ are both finite, then necessarily $\alpha(T) \leqslant \delta(T)$.*

PROOF. Let $p = \delta(T)$. Then (5.41–1) shows that

$$\mathscr{R}(T^p) = T\{\mathscr{R}(T^p) \cap \mathscr{D}(T)\}. \tag{5.41–2}$$

Now define a space X_1 and an operator T_1 as follows: $X_1 = \mathscr{R}(T^p)$, $\mathscr{D}(T_1) = \mathscr{R}(T^p) \cap \mathscr{D}(T)$, $T_1 x = Tx$ if $x \in \mathscr{D}(T_1)$. Equation (5.41–2) shows that $\mathscr{R}(T_1) = X_1$, so that $\delta(T_1) = 0$. It is easy to see that $\mathscr{D}(T_1{}^n) = \mathscr{R}(T^p) \cap \mathscr{D}_n(T)$ and $T_1{}^n x = T^n x$ if $x \in \mathscr{D}(T_1{}^n)$. This may be done by induction, using the fact that $\mathscr{R}(T^p) = \mathscr{R}(T^{p+1})$. Now clearly $\mathscr{N}(T_1^{n+1}) \subset \mathscr{N}(T^{n+1})$, and $\mathscr{N}(T^n) \cap \mathscr{D}(T_1^{n+1}) \subset \mathscr{N}(T^n) \cap \mathscr{D}(T_1{}^n) \subset \mathscr{N}(T_1{}^n)$.
It therefore follows that

$$\mathscr{N}(T_1^{n+1}) - \mathscr{N}(T_1{}^n) \subset \mathscr{N}(T^{n+1}) - \mathscr{N}(T^n).$$

Consequently $\alpha(T_1) \leqslant \alpha(T)$. Since $\alpha(T)$ is finite, so is $\alpha(T_1)$. But then $\alpha(T_1) = 0$, by Theorem 5.41–C, and so $T_1{}^{-1}$ exists. Now suppose that $x \in \mathscr{N}(T^{p+1})$, and let $y = T^p x$. Then $y \in \mathscr{R}(T^p) \cap \mathscr{D}(T) = \mathscr{D}(T_1)$. But $T_1 y = Ty = T^{p+1}x = 0$. Since $T_1{}^{-1}$ exists we conclude that $y = 0$, or $x \in \mathscr{N}(T^p)$. We thus have $\mathscr{N}(T^{p+1}) \subset \mathscr{N}(T^p)$, whence $\mathscr{N}(T^{p+1}) = \mathscr{N}(T^p)$. This means that $\alpha(T) \leqslant p = \delta(T)$, so the proof is complete.

Theorem 5.41–E. *If $\mathscr{D}(T) = X$ and the ascent and descent of T are both finite, they are equal.*

PROOF. We have only to prove $\alpha(T) \geqslant \delta(T)$, because of Theorem 5.41–D. The case $\delta(T) = 0$ is covered by Theorem 5.41–C; therefore we assume $p = \delta(T) \geqslant 1$. Then there exists an element $y \in \mathscr{R}(T^{p-1}) - \mathscr{R}(T^p)$. We can write $y = T^{p-1}x$ for some x. Let $z = Ty = T^p x$, so that $z \in \mathscr{R}(T^p)$. Now $T^p\{\mathscr{R}(T^p)\} = \mathscr{R}(T^{2p}) = \mathscr{R}(T^p)$. Therefore, there exists some $u \in \mathscr{R}(T^p)$ such that $T^p u = z$. Let $v = x - u$. Then

$$T^p v = T^p x - T^p u = z - z = 0,$$

and

$$T^{p-1}v = T^{p-1}x - T^{p-1}u = y - T^{p-1}u.$$

Now $T^{p-1}u \in \mathscr{R}(T^{2p-1}) = \mathscr{R}(T^p)$, because $u \in \mathscr{R}(T^p)$. Therefore $y - T^{p-1}u \neq 0$, because y is not in $\mathscr{R}(T^p)$. We see then that $v \in \mathscr{N}(T^p) - \mathscr{N}(T^{p-1})$, and this implies $\alpha(T) \geqslant p = \delta(T)$.

Our main interest in the ascent and descent of an operator arises from the facts set forth in the next two theorems.

Theorem 5.41–F. *Suppose that $\alpha(T)$ and $\delta(T)$ are both finite, and let $p = \delta(T)$. Then $\mathscr{R}(T^p)$ and $\mathscr{N}(T^p)$ are linearly independent linear manifolds, and*

$$\mathscr{D}_p(T) = \{\mathscr{R}(T^p) \cap \mathscr{D}_p(T)\} \oplus \mathscr{N}(T^p). \tag{5.41–3}$$

PROOF. Suppose $y \in \mathscr{R}(T^p) \cap \mathscr{N}(T^p)$. We can write $y = T^p x$, $T^p y = 0$, so that $T^{2p}x = 0$. Then $x \in \mathscr{N}(T^{2p})$. But $\alpha(T) \leqslant p$, and so $\mathscr{N}(T^{2p}) = \mathscr{N}(T^p)$. Hence $x \in \mathscr{N}(T^p)$, and $y = 0$. This proves that $\mathscr{R}(T^p) \cap \mathscr{N}(T^p) = (0)$, as required. For the proof of (5.41–3) we define X_1 and T_1 as in the proof of Theorem 5.41–D. Since $\mathscr{R}(T_1) = X_1$, it follows by Theorem 5.41–B that $\mathscr{R}(T_1{}^p) = X_1$. Hence, if $x \in \mathscr{D}_p(T)$, there is some $x_1 \in \mathscr{D}(T_1{}^p) = \mathscr{R}(T^p) \cap \mathscr{D}_p(T)$ such that $T_1{}^p x_1 = T^p x_1 = T^p x$. Then $x - x_1 \in \mathscr{N}(T^p)$. Let $x_2 = x - x_1$. Then $x = x_1 + x_2$ shows that

$$\mathscr{D}_p(T) \subset \{\mathscr{R}(T^p) \cap \mathscr{D}_p(T)\} \oplus \mathscr{N}(T^p).$$

The reverse inclusion is evident, and so (5.41–3) is proved.

Theorem 5.41–G. *Suppose that $\mathscr{D}(T) = X$ and that $\alpha(T)$ and $\delta(T)$ are both finite (and hence equal). Let $p = \delta(T)$. Then T is completely reduced by the pair of manifolds $\mathscr{R}(T^p)$, $\mathscr{N}(T^p)$, and T maps $\mathscr{R}(T^p)$ in a one-to-one manner onto all of itself.*

PROOF. The fact that $\mathscr{R}(T^p) \cap \mathscr{N}(T^p) = (0)$ and

$$X = \mathscr{R}(T^p) \oplus \mathscr{N}(T^p) \tag{5.41–4}$$

is known from Theorem 5.41–F, since $\mathscr{D}_p(T) = X$ in this case. The invariance of $\mathscr{N}(T^p)$ under T is obvious; that of $\mathscr{R}(T^p)$ follows from (5.41–2). The restriction of T to $\mathscr{R}(T^p)$ is T_1, as defined in the proof of Theorem 5.41–D. The asserted property of T_1 was established in this previous argument.

We use the results of this section in the study of compact operators, in § 5.5.

PROBLEMS

1. It can happen that $\delta(T) = 0$ and $\alpha(T) = \infty$. Consider $X = l^2$, $\mathscr{D}(T) = X$, and $x = (\xi_1, \xi_2, \cdots)$, $Tx = (\xi_2, \xi_3, \cdots)$.

2. If $\mathscr{D}(T) \neq X$ it can happen that $\alpha(T) < \delta(T)$ when both are finite. Take $\mathscr{D}(T) \neq X$, $Tx = x$ if $x \in \mathscr{D}(T)$. Then $\alpha(T) = 0$, $\delta(T) = 1$.

3. Suppose that X is a Banach space and T is a closed linear operator with domain and range in X. Then points λ of $\rho(T)$ are characterized by $\alpha(\lambda - T) = \delta(\lambda - T) = 0$; points of $P\sigma(T)$ by $\alpha(\lambda - T) \neq 0$; and points of $C\sigma(T) \cup R\sigma(T)$ by $\alpha(\lambda - T) = 0$ and $\delta(\lambda - T) > 0$.

5.5 Compact Operators

In our study of compact operators we shall prove theorems which, when applied to Fredholm integral equations of the second kind, yield a great part of the most fundamental information about such equations.

Definition. Let X and Y be normed linear spaces. Suppose T is a linear operator with domain X and range in Y. We say that T is *compact* if, for each bounded sequence $\{x_n\}$ in X, the sequence $\{Tx_n\}$ contains a subsequence converging to some limit in Y.

A compact operator is also called *completely continuous* (in French, *complètement* or *totalement continue*; in German, *vollstetig*). We note at once that a compact linear operator is continuous. For, discontinuity of T would imply the existence of a sequence $\{x_n\}$ such that $\|x_n\| \leqslant 1$ and $\|Tx_n\| \to \infty$, and this cannot occur if T is compact.

We call explicit attention to the important general facts about compact operators which are stated in problems 1–5 at the end of this section. Some other important facts are stated in the next two theorems.

Theorem 5.5–A. *If $T \in [X, Y]$ and T is compact, $\mathscr{R}(T)$ is separable.*

PROOF. Observe that $\mathscr{R}(T) = \bigcup_{n=1}^{\infty} T(S_n)$, where $S_n = \{x : \|x\| \leqslant n\}$. It will suffice to prove that $T(S_n)$ is separable. Now, since T is compact, it is easy to see that every infinite subset of $T(S_n)$ has an accumulation point in Y. Consequently (see the proof of Theorem 2.4–E) for each postive integer k there exists a finite set of points in $T(S_n)$ such that the spheres of radius $1/k$ with centers at these points cover $T(S_n)$. The aggregate of such finite sets, one set for each k, is evidently everywhere dense in $T(S_n)$.

Theorem 5.5–B. *If $T \in [X, Y]$ and T is compact, the conjugate operator T' is compact also.*

PROOF. Let $\{y_n\}$ be a sequence everywhere dense in $\mathscr{R}(T)$, and let $\{y_k'\}$ be a bounded sequence in Y', say with $\|y_k'\| \leqslant M$. Since the numerical sequence $y_1'(y_n), y_2'(y_n), \cdots$ is bounded for each n, we can use the diagonal procedure to obtain a subsequence of $\{y_k'\}$, which we shall denote by $\{z_k'\}$, such that $\lim\limits_{k\to\infty} z_k'(y_n)$ exists for each n. It is then true that $\lim\limits_{k\to\infty} z_k'(y)$ exists for each y in $\overline{\mathscr{R}(T)}$; this is easily seen from the inequalities

$$\begin{aligned} |z_k'(y) - z_j'(y)| \leqslant |z_k'(y) - z_k'(y_n)| & \\ + |z_k'(y_n) - z_j'(y_n)| + |z_j'(y_n) - z_j'(y)| & \\ \leqslant (\|z_k'\| + \|z_j'\|)\, \|y - y_n\| + |z_k'(y_n) - z_j'(y_n)|&, \end{aligned}$$

since $\{y_n\}$ is everywhere dense in $\overline{\mathscr{R}(T)}$ as well as in $\mathscr{R}(T)$. Now let $x_k' = T'z_k'$. If $x \in X$, we have $x_k'(x) = z_k'(Tx)$. Therefore $\lim\limits_{k\to\infty} x_k'(x)$ exists for each $x \in X$, and the limit defines an element $x' \in X'$. Our proof will be complete if we show that $T'z_k' \to x'$. We assume this is not the case. Then for some $\epsilon > 0$ there will exist a subsequence of $\{z_k'\}$, which we shall denote by $\{w_k'\}$, such that $\|T'w_k' - x'\| \geqslant \epsilon$ for every k. Now $w_k'(Tx) - x'(x)$ is the value at x of the functional $T'w_k' - x'$. Hence there exists $x_k \in X$, with $\|x_k\| = 1$, such that

$$|w_k'(Tx_k) - x'(x_k)| \geqslant \tfrac{1}{2}\|T'w_k' - x'\| \geqslant \epsilon/2. \tag{5.5–1}$$

By the compactness of T, some subsequence of $\{Tx_k\}$ is convergent. Denote this subsequence by $\{Tu_i\}$, and let $\{v_i'\}$ be the corresponding subsequence of $\{w_k'\}$. Let $Tu_i \to v$. Then $v \in \overline{\mathscr{R}(T)}$. Consequently $\lim\limits_{k\to\infty} v_k'(v) = \alpha$ exists. Now $x'(u_i) = \lim\limits_{k\to\infty} z_k'(Tu_i)$, by definition. Hence also $x'(u_i) = \lim\limits_{k\to\infty} v_k'(Tu_i)$, for $\{v_k'\}$ is a subsequence of $\{z_k'\}$. Next,

$$|v_k'(v) - v_k'(Tu_i)| \leqslant \|v_k'\|\, \|v - Tu_i\| \leqslant M\|v - Tu_i\|,$$

and so

$$|\alpha - x'(u_i)| \leqslant M\|v - Tu_i\|. \tag{5.5–2}$$

Now $|v_i'(Tu_i) - x'(u_i)| \geqslant \epsilon/2$, in view of (5.5–1). But

$$\begin{aligned}|v_i'(Tu_i) - x'(u_i)| &\leqslant |v_i'(Tu_i) - v_i'(v)| + |v_i'(v) - \alpha| + |\alpha - x'(u_i)| \\ &\leqslant 2M\|Tu_i - v\| + |v_i'(v) - \alpha|,\end{aligned}$$

in view of (5.5–2). The right-hand member of this last inequality tends to 0 as $i \to \infty$, and so we reach a contradiction. The proof is then complete.

The converse of Theorem 5.5–B is valid if Y is complete. See problem 5.

Before going further with our general study of compact operators, let us consider some examples.

Example 1. Let $X = C[a, b]$, where $[a, b]$ is a finite closed interval. The scalars may be either real or complex. Let $k(s, t)$ be a continuous function of s and t on $[a, b] \times [a, b]$, and let $K \in [X]$ be defined as the Fredholm-type integral operator with kernel k [see (4.11–1)]. Then K is compact. The proof of this fact depends on a theorem known as Ascoli's theorem, which involves the notion of an *equicontinuous family*. A family F of elements of $C[a, b]$ is called equicontinuous if to each $\epsilon > 0$ corresponds a $\delta > 0$ such that $|x(t_1) - x(t_2)| < \epsilon$ whenever $x \in F$ and t_1, t_2 are points of $[a, b]$ such that $|t_1 - t_2| < \delta$.

Ascoli's Theorem. *If $\{x_n\}$ is a bounded sequence from $C[a, b]$ such that the x_n's form an equicontinuous family, then $\{x_n\}$ contains a convergent subsequence (convergent in the topology of $C[a, b]$).*

This is a standard theorem. For reference see Graves, **1**, page 122, or McShane, **1**, page 336.

Now consider any bounded sequence $\{x_n\}$ in $C[a, b]$, and let $y_n = Kx_n$. Clearly $\{y_n\}$ is bounded, for $\|y_n\| \leqslant \|K\| \|x_n\|$. Hence K will be proved compact if we can show that the y_n's form an equicontinuous family. Now $k(s, t)$ is uniformly continuous, and hence to each $\epsilon > 0$ there corresponds some $\delta > 0$ such that $|k(s_1, t) - k(s_2, t)| < \epsilon/(b - a)M$ for all t if $|s_1 - s_2| < \delta$; here M denotes $\sup_n \|x_n\|$. Then for each n we have

$$|y_n(s_1) - y_n(s_2)| = \left|\int_a^b [k(s_1, t) - k(s_2, t)]x_n(t)\, dt\right| \leqslant \epsilon$$

if $|s_1 - s_2| < \delta$; thus the proof is completed.

The operator K will still be compact with certain less severe restrictions on the kernel k. If $k(s, t)$ is of class $\mathscr{L}^2(a, b)$ as a function of t for each s, if

$$\int_a^b |k(s, t)|^2\, dt$$

is a bounded function of s, and, if

$$\int_a^b |k(s_1, t) - k(s_2, t)|^2 \, dt \to 0 \qquad \text{as} \qquad |s_1 - s_2| \to 0,$$

the operator K will be compact as an operator acting in $C[a, b]$. It is also compact as an operator acting in $L^2(a, b)$. The proof that K maps a bounded sequence into an equicontinuous sequence is made by using the Schwarz inequality.

Example 2. If $k(s, t)$ is an $\mathscr{L}^2$ kernel (see § 4.12), the corresponding integral operator K, acting in the space $L^2(a, b)$, is compact. We prove this as follows. Choose a complete orthonormal set $u_1, u_2, \cdots$ for $L^2(a, b)$. With the usual notation for inner products, let $\alpha_{ij} = (Ku_j, u_i)$, and let $v_j = Ku_j$. Since $\overline{k(s, t)}$ is of class $\mathscr{L}^2$ as a function of t for almost all s, we have

$$\int_a^b |k(s, t)|^2 \, dt =^{0} \sum_{j=1}^{\infty} \left| \int_a^b \overline{k(s, t)} \, \overline{u_j(t)} \, dt \right|^2 = \sum_{j=1}^{\infty} |v_j(s)|^2.$$

Integrating with respect to s, we get

$$\int_a^b \int_a^b |k(s, t)|^2 \, dt \, ds = \sum_{j=1}^{\infty} \|Ku_j\|^2.$$

But

$$\|Ku_j\|^2 = \sum_{i=1}^{\infty} |(Ku_j, u_i)|^2 = \sum_{i=1}^{\infty} |\alpha_{ij}|^2.$$

Therefore

$$\int_a^b \int_a^b |k(s, t)|^2 \, ds \, dt = \sum_{i,j=1}^{\infty} |\alpha_{ij}|^2. \tag{5.5–3}$$

Now define an operator K_n by the formula

$$K_n x = \sum_{i,j=1}^{n} \alpha_{ij}(x, u_j)u_i, \qquad x \in L^2(a, b).$$

This operator has the $\mathscr{L}^2$ kernel

$$k_n(s, t) = \sum_{i,j=1}^{n} \alpha_{ij} u_i(s) \overline{u_j(t)}.$$

It is easily verified that

$$(K_n u_j, u_i) = \begin{cases} \alpha_{ij} & \text{if } i \leqslant n \text{ and } j \leqslant n \\ 0 & \text{if } i \text{ or } j > n. \end{cases}$$

Hence, by (5.5–3) applied to $K - K_n$ instead of K, we have

$$\int_a^b \int_a^b |k(s, t) - k_n(s, t)|^2 \, ds \, dt = \sum_{i,j=1}^{\infty} |\alpha_{ij}|^2 - \sum_{i,j=1}^{n} |\alpha_{ij}|^2 \to 0$$

as $n \to \infty$. But, as we know from § 4.12,

$$\|K - K_n\|^2 \leqslant \int_a^b \int_a^b |k(s, t) - k_n(s, t)|^2 \, ds \, dt.$$

Hence $\|K - K_n\| \to 0$. Now, K_n is compact, for its range is finite dimensional (see problem 2). The fact that $K_n \to K$ then assures us that K is compact (see problem 3).

Example 3. If $1 \leqslant q < \infty$, a continuous linear operator A on l^1 into l^q is represented by an infinite matrix (α_{ij}), where the condition on the matrix is that

$$\|A\| = \sup_j \left(\sum_{i=1}^{\infty} |\alpha_{ij}|^q \right)^{1/q}.$$

be finite (see problem 5, § 4.51). The operator will be compact if and only if, in addition,

$$\sum_{i=n}^{\infty} |\alpha_{ij}|^q \to 0 \qquad \text{as } n \to \infty, \text{ uniformly in } j. \tag{5.5–4}$$

That this is a sufficient condition for compactness of A may be seen by using the result of problem 7, for, if $y = Ax$ it follows from Minkowski's inequality that

$$\left(\sum_{i=n}^{\infty} |\eta_i|^q \right)^{1/q} \leqslant \|x\| \left(\sum_{i=n}^{\infty} |\alpha_{ij}|^q \right)^{1/q}.$$

The condition (5.5–4) is also necessary, again by problem 7, for the vectors $v_j = (\alpha_{1j}, \alpha_{2j}, \cdots)$ must form a conditionally sequentially compact set, owing to the compactness of A (note that $v_j = Au_j$, where u_j is the vector whose i^{th} component is δ_{ij}).

We now turn to the study of a compact linear operator T which belongs to $[X]$. We start by investigating the null manifold and range of $(\lambda - T)^n$ when $\lambda \neq 0$. We do not require that X be complete. For convenience we write $T_\lambda = \lambda - T$.

Theorem 5.5–C. *Suppose $T \in [X]$, T compact, and $\lambda \neq 0$. Then the null manifolds $\mathcal{N}(T_\lambda{}^n)(n = 1, 2, \cdots)$ are finite dimensional.*

PROOF. We begin with $n = 1$. By Theorem 3.12–F it suffices to show that the set $\{x : \|x\| = 1\} \cap \mathcal{N}(T_\lambda)$ is compact or that every sequence from it contains a convergent subsequence. Suppose $\|x_n\| = 1$ and $x_n \in \mathscr{N}(T_\lambda)$, so that $x_n = \lambda^{-1}Tx_n$. Since T is compact it is now evident that $\{x_n\}$ contains a convergent subsequence. For $n > 1$ we write

$$T_\lambda{}^n = (\lambda - T)^n = \lambda^n - n\lambda^{n-1}T + \cdots + (-1)^nT^n = \lambda^n - TA,$$

where A is a certain member of $[X]$. Now TA is compact, since T is (see problem 4). The foregoing reasoning, applied to $\lambda^n - TA$, shows that $\mathscr{N}(T_\lambda{}^n)$ is finite dimensional.

Theorem 5.5–D. *Suppose $T \in [X]$, T compact, and $\lambda \neq 0$. Then the ranges $\mathscr{R}(T_\lambda{}^n)$ are all closed.*

PROOF. Just as in the foregoing proof it suffices to treat the case $n = 1$. We suppose that $\mathscr{R}(T_\lambda)$ is not closed; hence there is a sequence $\{T_\lambda x_n\}$ converging to a limit y, with y not in $\mathscr{R}(T_\lambda)$. Then $y \neq 0$, and hence x_n is not in $\mathscr{N}(T_\lambda)$ if n is sufficiently large. We may therefore suppose that no x_n is in $\mathscr{N}(T_\lambda)$. Since $\mathscr{N}(T_\lambda)$ is closed, the distance d_n from x_n to $\mathscr{N}(T_\lambda)$ is positive. Choose $u_n \in \mathscr{N}(T_\lambda)$ so that $\|x_n - u_n\| < 2d_n$. Let $\theta_n = \|x_n - u_n\|$. We shall prove that $\theta_n \to \infty$. If this were not so, $\{x_n - u_n\}$ would contain a bounded subsequence, and hence $T(x_n - u_n)$ would contain a convergent subsequence. But

$$x_n - u_n = (\lambda^{-1})\,[T_\lambda(x_n - u_n) + T(x_n - u_n)],$$

and $T_\lambda(x_n - u_n) = T_\lambda x_n \to y$, so that $\{x_n - u_n\}$ would also contain a convergent subsequence with some limit x. The corresponding subsequence of $\{T_\lambda(x_n - u_n)\}$ would then converge both to $T_\lambda x$ and to y, whence $y = T_\lambda x \in \mathscr{R}(T)$, contrary to assumption. Thus $\theta_n \to \infty$. Now let $v_n = (x_n - u_n)/\theta_n$. Observe that

$$T_\lambda v_n = (1/\theta_n)T_\lambda x_n \to 0. \tag{5.5–5}$$

Now

$$v_n = (\lambda^{-1})[T_\lambda v_n + Tv_n]$$

and $\|v_n\| = 1$. From (5.5–5) and the compactness of T it follows that $\{v_n\}$ contains a convergent subsequence with some limit v. From (5.5–5) it follows that $T_\lambda v = 0$, i.e., $v \in \mathscr{N}(T_\lambda)$. Now let $w_n = u_n + \theta_n v$. Observe that $w_n \in \mathscr{N}(T_\lambda)$, so that $d_n \leqslant \|x_n - w_n\|$. On the other hand, $x_n - w_n = \theta_n(v_n - v)$, so that $\|x_n - w_n\| < 2d_n\|v_n - v\|$, since $\theta_n < 2d_n$. We see in this way that $1 < 2\|v_n - v\|$, which contradicts the fact that $\{v_n\}$ contains a subsequence converging to v. We must therefore conclude that $\mathscr{R}(T_\lambda)$ is closed.

Theorem 5.5–E. *Suppose $T \in [X]$, T compact, and $\lambda \neq 0$. Then the ascent and descent of $\lambda - T$ are both finite (and hence equal).*

PROOF. Suppose $\alpha(\lambda - T) = \infty$. Then $\mathcal{N}(T^{n-1})$ is a proper closed subset of $\mathcal{N}(T_\lambda{}^n)$ for $n = 1, 2, \cdots$. By Theorem 3.12–E there exists $x_n \in \mathcal{N}(T_\lambda{}^n)$ such that $\|x_n\| = 1$ and $\|x_n - x\| \geqslant 1/2$ if $x \in \mathcal{N}(T^{n-1})$. Assume $1 \leqslant m < n$ and let

$$z = x_m + \lambda^{-1}T_\lambda x_n - \lambda^{-1}T_\lambda x_m.$$

Then

$$T_\lambda^{n-1}z = T_\lambda^{n-1}x_m + \lambda^{-1}T_\lambda^n x_n - \lambda^{-1}T_\lambda^n x_m = 0,$$

and so $z \in \mathcal{N}(T_\lambda^{n-1})$. Consequently $\|x_n - z\| \geqslant 1/2$. But we easily calculate that $Tx_n - Tx_m = \lambda(x_n - z)$, and so $\|Tx_n - Tx_m\| \geqslant |\lambda|/2 > 0$. This shows that $\{Tx_n\}$ can have no convergent subsequence,in contradiction to the fact that T is compact. Thus $\alpha(\lambda - T)$ must be finite.

The proof that $\delta(\lambda - T)$ is finite is similar. If it were not, $\mathcal{R}(T_\lambda{}^n)$ would be a proper closed subset of $\mathcal{R}(T_\lambda^{n-1})$ for $n = 1, 2, \cdots$. We choose $y_n \in \mathcal{R}(T_\lambda^n)$ so that $\|y_n\| = 1$ and $\|y_n - y\| \geqslant 1/2$ if $y \in \mathcal{R}(T_\lambda^{n+1})$. If $1 \leqslant m < n$ let

$$w = y_n - \lambda^{-1}T_\lambda y_n + \lambda^{-1}T_\lambda y_m.$$

We can write $y_k = T^k x_k$ for some x_k. Thus

$$w = T_\lambda^n x_n - \lambda^{-1}T_\lambda^{n+1}x_n + \lambda^{-1}T_\lambda^{m+1}x_m,$$

from which we see that $w \in \mathcal{R}(T_\lambda^{m+1})$. Therefore $\|y_m - w\| \geqslant 1/2$. But $Ty_m - Ty_n = \lambda(y_m - w)$ and so $\|Ty_m - Ty_n\| \geqslant |\lambda|/2$, and we obtain a contradiction, just as in the earlier argument. This completes the proof.

Our next main objective is to show that when $\lambda \neq 0$ and T is compact, λ is either an eigenvalue of T or in the resolvent set of T. For this purpose we prove that there exists a constant M such that if $d(x)$ is the distance from x to $\mathcal{N}(T_\lambda)$ (x arbitrary), then

$$d(x) \leqslant M\|T_\lambda x\|. \tag{5.5–6}$$

If we suppose this not true, there exists a sequence $\{x_n\}$ of points not in $\mathcal{N}(T_\lambda)$ such that $d(x_n)/\|T_\lambda x_n\| \to \infty$. Since $\mathcal{N}(T_\lambda)$ is finite dimensional, there exists $u_n \in \mathcal{N}(T_\lambda)$ such that $\|x_n - u_n\| = d(x_n)$. Let $y_n = (x_n - u_n)/d(x_n)$. Then $\|y_n\| = 1$ and $T_\lambda y_n \to 0$. From the equation

$$y_n = \lambda^{-1}(T_\lambda y_n + Ty_n)$$

and the compactness of T we infer that $\{y_n\}$ contains a subsequence converging to a limit y. Then the fact that $T_\lambda y_n \to 0$ implies $T_\lambda y = 0$. Then $u_n + d(x_n)y \in \mathcal{N}(T_\lambda)$, and so

$$\|y_n - y\| = \frac{\|x_n - [u_n + d(x_n)y]\|}{d(x_n)} \geqslant 1.$$

This contradicts the fact that $\{y_n\}$ contains a subsequence converging to y. Thus (5.5–6) is proved.

Theorem 5.5–F. *Suppose $T \in [X]$, T compact, and $\lambda \neq 0$. If M is the constant in (5.5–6), then, to each $y \in \mathscr{R}(T_\lambda)$ there corresponds an x such that $T_\lambda x = y$ and $\|x\| \leqslant M\|y\|$. Hence $(\lambda - T)^{-1}$ is continuous if it exists. The point λ is either in $\rho(T)$, in which case $\mathscr{R}(T_\lambda) = X$, or it is in $P\sigma(T)$, in which case $\mathscr{R}(T_\lambda)$ is a proper closed subspace of X.*

PROOF. To prove the first assertion, suppose $y = T_\lambda x_1$. Choose $u \in \mathscr{N}(T_\lambda)$ so that $\|x_1 - u\| = d(x_1)$, and let $x = x_1 - u$. Then $T_\lambda x = y$ and $\|x\| = d(x_1) \leqslant M\|T_\lambda x_1\| = M\|y\|$. We know that either $\alpha(\lambda - T) = \delta(\lambda - T) = 0$ or $0 < \alpha(\lambda - T) = \delta(\lambda - T) < \infty$. In the first case $(\lambda - T)^{-1}$ exists and $\mathscr{R}(T_\lambda) = X$, as we know from the definitions of ascent and descent. In the second case $(\lambda - T)^{-1}$ does not exist and $\mathscr{R}(T_\lambda) \neq X$. The rest of the assertions are evident from what has been shown.

The next question which arises naturally is this: How many points can there be in $P\sigma(T)$? This question is answered in the next theorem.

Theorem 5.5–G. *Suppose $T \in [X]$ and T compact. Then $P\sigma(T)$ contains at most a countable set of points, and these have no accumulation point except possibly the point $\lambda = 0$.*

PROOF. We first observe the following: If $Tx_k = \lambda_k x_k (k = 1, \cdots, n)$, where $x_k \neq 0$ and the λ_k's are all distinct, then $x_1, \cdots, x_n$ are linearly independent. For, if not, let x_m be the first x_k which is a linear combination of its predecessors, say $x_m = \alpha_1 x_1 + \cdots + \alpha_{m-1}x_{m-1}$. From $Tx_m - \lambda_m x_m = 0$ we obtain

$$0 = \alpha_1(\lambda_1 - \lambda_m)x_1 + \cdots + \alpha_{m-1}(\lambda_{m-1} - \lambda_m)x_{m-1}.$$

Hence $\alpha_1 = \cdots = \alpha_{m-1} = 0$, whence $x_m = 0$, which is a contradiction.

Now, to prove the theorem it suffices to show that, if $\epsilon > 0$, the points λ of $P\sigma(T)$ for which $|\lambda| \geqslant \epsilon$ form a finite set. Suppose the contrary for some ϵ, and let $\{\lambda_n\}$ be an infinite sequence of distinct points in $P\sigma(T)$ with $|\lambda_n| \geqslant \epsilon$. Then $Tx_n = \lambda_n x_n$, where $x_n \neq 0$. The set of all the x_n's is linearly independent. Let M_n be the linear manifold generated by $x_1, \cdots, x_n$. Then M_{n-1} is a proper closed subset of M_n. By Theorem 3.12–E there exists $u_n \in M_n$ such that $\|u_n\| = 1$ and $\|u_n - x\| \geqslant 1/2$ if $x \in M_{n-1}$. Now $Tu_n \in M_n$, since $Tx_k = \lambda_k x_k$ and u_n is a linear combination of $x_1, \cdots, x_n$. But if $x = \alpha_1 x_1 + \cdots + \alpha_n x_n$ we see that

$$(\lambda_n - T)x = \alpha_1(\lambda_n - \lambda_1)x_1 + \cdots + \alpha_{n-1}(\lambda_n - \lambda_{n-1})x_{n-1},$$

so that $(\lambda_n - T)(M_n) \subset M_{n-1}$. In particular, $(\lambda_n - T)u_n \in M_{n-1}$. Therefore $z = (\lambda_n - T)u_n + Tu_m \in M_{n-1}$ if $1 \leqslant m < n$, and so also $\lambda_n^{-1}z \in M_{n-1}$. Now

$$Tu_n - Tu_m = \lambda_n u_n - (\lambda_n u_n - Tu_n + Tu_m) = \lambda_n(u_n - \lambda_n^{-1}z),$$

and so $\|Tu_n - Tu_m\| \geqslant |\lambda_n|/2 \geqslant \epsilon/2$. This contradicts the fact that $\{Tu_n\}$ must contain a convergent subsequence and completes the proof of the theorem.

By various examples it may be shown that when T is compact the point $\lambda = 0$ can belong to $P\sigma(T)$, $C\sigma(T)$, or $R\sigma(T)$; it cannot belong to $\rho(T)$ if X is infinite dimensional. See problems 9–12.

The next phase of our study concerns the conjugate operator T'. We know that it is compact (Theorem 5.5–B). From Theorem 5.3–A we know that T and T' have the same spectrum. Hence, if $\lambda \neq 0$, then λ is an eigenvalue of T if and only if it is an eigenvalue of T'. But more than this is true.

Theorem 5.5–H. *Suppose $T \in [X]$, T compact, and $\lambda \neq 0$. Then for each n the null spaces of T_λ^n and $(T_\lambda')^n$ have the same dimension, and hence $\lambda - T$ and $\lambda - T'$ have the same ascent.*

PROOF. It suffices to give the proof for $n = 1$, because $(T_\lambda^n)' = (T_\lambda')^n$ and $T_\lambda^n = \lambda^n - B$, where B is compact (see the proof of Theorem 5.5–C). Suppose $\mathscr{N}(T_\lambda)$ and $\mathscr{N}(T_\lambda')$ are of dimensions m and n, respectively. We can assume $m \geqslant 1$, $n \geqslant 1$, because $m = 0$ implies $n = 0$ and conversely. Let $x_1, \cdots, x_m$ and $u_1', \cdots, u_n'$ be bases for $\mathscr{N}(T_\lambda)$ and $\mathscr{N}(T_\lambda')$, respectively. It follows easily from Theorem 3.5–C that there exists a set of elements $u_1, \cdots, u_n$ in X such that

$$u_i'(u_j) = \delta_{ij}. \tag{5.5–7}$$

Also, there is a set of elements $x_1', \cdots, x_m'$ in X' such that

$$x_i'(x_j) = \delta_{ij}. \tag{5.5–8}$$

This follows from Theorem 3.8–E (or from Theorem 4.3–D).

We assume $m < n$ and deduce a contradiction. We define an operator A by the equation

$$Ax = Tx + \sum_{i=1}^{m} x_i'(x)u_i;$$

A is compact, by problems 2 and 3. Applying Theorem 4.6–C to T_λ, we see that

$$u_j'(T_\lambda x) = 0, \qquad x \in X, \qquad j = 1, \cdots, n. \tag{5.5–9}$$

We now prove that $(\lambda - A)^{-1}$ exists. For,

$$(5.5\text{–}10) \qquad (\lambda - A)x = T_\lambda x - \sum_{i=1}^{m} x_i'(x)u_i.$$

If $(\lambda - A)x = 0$, we conclude from (5.5–10), (5.5–9) and (5.5–8) that $x_1'(x) = \cdots x_m'(x) = 0$, whence $T_\lambda x = 0$. But then x is a linear combination of $x_1, \cdots, x_m$, say $x = \alpha_1 x_1 + \cdots + \alpha_m x_m$. Applying x_i' to x, we find that $\alpha_i = 0$, whence $x = 0$. Thus $(\lambda - A)^{-1}$ exists. By Theorem 5.5–F, applied to A, we conclude that $\mathscr{R}(\lambda - A) = X$. We then choose x so that $(\lambda - A)x = u_{m+1}$. Now $u'_{m+1}(u_{m+1}) = 1$. But by (5.5–10), (5.5–9) and (5.5–7) we have

$$u'_{m+1}(u_{m+1}) = u'_{m+1}(T_\lambda x) - \sum_{i=1}^{m} x_i'(x)u'_{m+1}(u_i) = 0.$$

From this contradiction we must conclude that $m \geqslant n$.

Next we assume that $m > n$ and deduce a contradiction. We define an operator B on X' by

$$Bx' = T'x' + \sum_{i=1}^{n} x'(u_i)x_i',$$

and observe that it is compact. This time we use Theorem 4.6–F to observe that $x_i \in {}^0\{\mathscr{R}(T_\lambda')\}$, and from this we prove that $(\lambda - B)^{-1}$ exists. We then choose x' so that $(\lambda - B)x' = x'_{n+1}$ and obtain a contradiction of the fact that $x'_{n+1}(x_{n+1}) = 1$. The reader can easily supply the details. Thus, finally, we conclude that $m = n$.

The last of the theorems of this section tells us the exact relation between $\mathscr{R}(T_\lambda)$ and $\mathscr{N}(T_\lambda')$, with a dual result when T_λ and T_λ' are exchanged.

Theorem 5.5–I. *Suppose* $T \in [X]$, T *compact, and* $\lambda \neq 0$. *Then* $\mathscr{R}(T_\lambda) = {}^0\{\mathscr{N}(T_\lambda')\}$, *and* $\mathscr{R}(T_\lambda') = \{\mathscr{N}(T_\lambda)\}^0$.

PROOF. The first relation follows from Theorem 4.6–D, because we know that $\mathscr{R}(T_\lambda)$ is closed. For the second relation we have only to prove that $\{\mathscr{N}(T_\lambda)\}^0 \subset \mathscr{R}(T_\lambda')$, because of Theorem 4.6–G. Hence we start by assuming $y' \in \{\mathscr{N}(T_\lambda)\}^0$. If $y \in \mathscr{R}(T_\lambda)$, define $f(y) = y'(x)$, where x is any element such that $T_\lambda x = y$. The value of $y'(x)$ is the same, no matter which x is chosen, because $y' \in \{\mathscr{N}(T_\lambda)\}^0$. Clearly f is a linear functional on $\mathscr{R}(T_\lambda)$. Now, with y given, we can choose x so that $T_\lambda x = y$ and $\|x\| \leqslant M\,\|y\|$ (Theorem 5.5–F). From this it follows easily that f is continuous. If we now choose $x' \in X'$ so that x' is an extension of f

(and we can do this), we have $x'(T_\lambda x) = y'(x)$ for every $x \in X$, and so $y' = T_\lambda' x'$. Thus $y' \in \mathscr{R}(T_\lambda')$, and the proof is complete.

We conclude this section with a brief summary of the application of the theory of compact operators to the study of integral equations. We consider Fredholm equations of the second kind, in the form

$$y(s) = x(s) - \lambda \int_a^b k(s, t)x(t)\, dt, \tag{5.5–11}$$

where the kernel k is of such a sort that the integral operator K defined by it is compact when acting in a certain specified normed linear function space X. We think primarily of $X = C[a, b]$ or $X = L^2(a, b)$. We can write (5.5–11) in the form $y = x - \lambda Kx$. Note that the parameter λ here occupies a different position than it does in our previous work, where we have been studying the equation $y = \lambda x - Tx$. There is no interest in the equation (5.5–11) when $\lambda = 0$, so we can assume $\lambda \neq 0$ and rewrite the equation as

$$\lambda^{-1} y = \lambda^{-1} x - Kx.$$

Since $\lambda^{-1} \neq 0$, our theory of compact operators now tells us the following:

For a given $\lambda \neq 0$, *either* (5.5–11) *has a unique solution* x *corresponding to each choice of* $y \in X$, *or else the homogeneous equation*

$$0 = x(s) - \lambda \int_a^b k(s, t)x(t)\, dt \tag{5.5–12}$$

has a solution $x \neq 0$. *In this latter case the number of linearly independent solutions of* (5.5–12) *is finite.* The number λ^{-1} is then an eigenvalue of the integral operator K. It is customary to call λ itself a *characteristic value* of K. *The set of characteristic values is at most countable. If there is an infinite sequence* $\{\lambda_n\}$ *of such values,* $|\lambda_n| \to \infty$.

If the function space in question is $C[a, b]$, the conjugate operator K' acts in a different space. But, if we are dealing with $L^2(a, b)$ and if k is an $\mathscr{L}^2$ kernel, we can identify K' with the integral operator in $L^2(a, b)$ such that $K'x$ is represented by

$$\int_a^b k(t, s)x(t)\, dt.$$

In this case the theory goes on to tell us that (5.5–12) *and the equation*

$$0 = x(s) - \lambda \int_a^b k(t, s)x(t)\, dt \tag{5.5–13}$$

have the same number of linearly independent solutions. Furthermore, if

λ is a characteristic value, (5.5–11) *has a solution in L^2 corresponding to a given $y \in L^2$ if and only if*

$$\int_a^b y(t)\overline{x(t)}\, dt = 0$$

whenever x is a solution of (5.5–13).

Even when we are interested in the space $C[a, b]$ the L^2 theory may be useful to us. Suppose that k is an $\mathscr{L}^2$ kernel such that Kx and $K'x$ are in $C[a, b]$ whenever $x \in L^2(a, b)$. This is the case, for instance, if $k(s, t)$ is continuous in both variables (we assume that $[a, b]$ is a finite interval). Then, if $y \in C[a, b]$ and $x \in L^2(a, b)$ is a solution of (5.5–11), it follows that $x \in C[a, b]$. Hence, in particular, the solutions of (5.5–12) are continuous, and the same is true of (5.5–13). Therefore it follows that all of the foregoing italicized assertions remain valid if we regard the $\mathscr{L}^2$ kernel as defining an operator in $C[a, b]$ rather than in $L^2(a, b)$ and restrict x and y accordingly.

The theory of Fredholm integral equations of the second kind, in the form here presented, can be applied to obtain a solution of the Dirichlet problem in two dimensions, provided the region in question is of a sufficiently simple character (the essential restrictions are on the smoothness of the boundary). The same method can be applied to other boundary-value problems associated with the Laplace operator, e.g., the Neumann problem. For the Dirichlet problem in spaces of higher dimension, the procedure is slightly more complicated, because one has to deal with an integral operator which may not be compact. However, a certain power of the operator turns out to be compact, and this is sufficient to obtain the results stated in the first two sentences in connection with (5.5–12) (see the remarks following the proof of Theorem 5.8–F). These are the decisive results for the application to the Dirichlet problem. For more details see Riesz and Nagy, **1**, pages 189–192 and Kellogg, **1**, pages 311–315.

PROBLEMS

1. If X is finite dimensional and T is linear on X into Y, then T is compact.

2. If $T \in [X, Y]$ and Y is finite dimensional, then T is compact.

3. The set of compact linear operators on X into Y is a linear manifold in $[X, Y]$. If Y is complete, this manifold is closed. For proof of the last assertion assume A_k compact and in $[X, Y]$, $B \in [X, Y]$, and $\|A_k - B\| \to 0$. If $\{x_n\}$ is a bounded sequence in X, use the diagonal procedure to obtain a subsequence of $\{x_n\}$, denoted by $\{u_n\}$, such that $\lim_{n\to\infty} A_k(u_n)$ exists for each k. Then show that $\{Bu_n\}$ is a Cauchy sequence in Y.

4. If A, $B \in [X]$ and A is compact, so are AB and BA.

5. If $T \in [X, Y]$, Y is complete, and T' is compact, then T is compact. For the proof use the relation $T''J_1 = J_2T$ (using notation as in problem 3, § 4.5) and the fact that T'' is compact.

6. If $X = l^2$ and (α_{ij}) is an infinite matrix such that $\sum_{i,j=1}^{\infty} |\alpha_{ij}|^2 < \infty$, the equations $\eta_i = \sum_{j=1}^{\infty} \alpha_{ij}\xi_j$ define a compact linear mapping $y = Ax$ of X into itself, with $\|A\| \leqslant \left(\sum_{i,j=1}^{\infty} |\alpha_{ij}|^2 \right)^{\frac{1}{2}}$.

7. A set S in a normed linear space X is called conditionally sequentially compact if every sequence from S contains a convergent subsequence. If $X = l^p$, where $1 \leqslant p < \infty$, a set S in X has this property if and only if (a) S is bounded, and (b) $\sum_{i=n}^{\infty} |\xi_i|^p \to 0$ as $n \to \infty$, uniformly for all $x = \{\xi_i\}$ in S. To prove the sufficiency of these conditions use a diagonal procedure on the components of the given sequence $\{x_n\}$. Boundedness of S is clearly necessary; the necessity of (b) may be proved by the method of contradiction.

8. Show from Theorem 5.5–F that $\lambda - T$ takes open sets in X into sets which are open in the induced topology of $\mathscr{R}(T_\lambda)$.

9. The point $\lambda = 0$ has a status different from other points in relation to T if $T \in [X]$, T is compact, and X is infinite dimensional. In this case 0 can never be in $\rho(T)$. To show this, assume the contrary, so that $\overline{\mathscr{R}(T)} = X$ and T has a continuous inverse. Suppose $\|y_n\| = 1$. Choose $z_n = Tx_n$ so that $\|z_n - y_n\| < 1/n$. Use the compactness of T to deduce that $\{y_n\}$ contains a convergent subsequence and hence that X is finite dimensional.

10. Define $T \in [l^1]$ by $y = \{k^{-1}\xi_k\}$, where $x = \{\xi_k\}$. Then T is compact, and $0 \in C\sigma(T)$. What are the eigenvalues?

11. Define $T \in [l^1]$ by the matrix (t_{ij}) with $t_{i1} = 2^{1-i}$, $i \geqslant 2$, $t_{ii} = 2^{1-i}$, $i \geqslant 2$, and all other $t_{ij} = 0$. Then T is compact, and $0 \in R\sigma(T)$. What are the eigenvalues?

12. Suppose $x_0 \in X$, $x_0' \in X'$, $x_0'(x_0) \neq 0$. Define $Tx = x_0'(x)x_0$. Show that $\sigma(T)$ consists of the eigenvalues $\lambda = 0$, $\lambda = x_0'(x_0)$ and that $R_\lambda y = \lambda^{-1}y + \lambda^{-1}[\lambda - x_0'(x_0)]^{-1}Ty$.

13. Suppose X is not complete, and let $T \in [X]$ be compact. Let $\hat{T}$ be the unique continuous extension of T to all of the completion $\hat{X}$. Then $\hat{T}$ is compact and $\mathscr{R}(\hat{T}) \subset X$. If $\lambda \neq 0$, T_λ^n and $\hat{T}_\lambda^n$ have the same null manifold, and so T_λ and $\hat{T}_\lambda$ have the same ascent; also, $\mathscr{R}(T_\lambda^n) = \mathscr{R}(\hat{T}_\lambda^n) \cap X$ and $\hat{T}_\lambda^n(\hat{X} - X) \subset \hat{X} - X$. Finally, $\sigma(T) = \sigma(\hat{T})$.

14. Suppose $f \in C[a, b]$. Consider the inhomogeneous two-point problem

$$y''(s) + \lambda f(s)y(s) = x(s), \qquad y(a) = \alpha, \qquad y(b) = \beta,$$

and the corresponding homogeneous problem in which the function x is identically zero. Show that, for a given λ, either the inhomogeneous problem has a unique twice continuously differentiable solution corresponding to each $x \in C[a, b]$, or else the homogeneous problem has a nonzero twice continuously differentiable solution. Also, the λ's for which the latter situation occurs form an at most countable set with no finite point of accumulation in the extended complex plane. Use the discussion in § 4.13, with $\lambda f(s)$ replacing $a_2(s)$.

15. Given $T \in [X, Y]$ (X and Y normed linear spaces) and T compact, then $Tx_n \to Tx$ whenever $\{x_n\}$ is X'-weakly convergent to x. The essential thing to show is that, if $\{x_n\}$ is X'-weakly convergent to x, a subsequence of $\{Tx_n\}$ converges to Tx. This can be done with the aid of the conjugate operator T'. One can then conclude that $Tx_n \to Tx$, by assuming this false and deducing a contradiction.

16. Suppose $T \in [X, Y]$, where X is norm reflexive and T has the property that $Tx_n \to Tx$ whenever $\{x_n\}$ is X'-weakly convergent to x. Then T is compact. For operators in Hilbert space, this is the form in which the concept of a *vollstetig* operator was originally introduced. The proposition does not remain true if the hypothesis that X is norm reflexive is dropped. Example: take $X = Y = l^1$. Then the identity mapping is not compact, whereas X'-weak convergence is identical with convergence according to the norm (Banach, **1**, page 137).

5.6 An Operational Calculus

If X is a complex Banach space and T is a closed operator with domain and range in X, the fact that the resolvent operator $R_\lambda \equiv (\lambda - T)^{-1}$ is analytic as a function of λ enables us to obtain some important results by the use of contour integrals in the complex λ-plane. Since X is complete, so is $[X]$; the completeness of these spaces assures us of the existence of the contour integrals with which we deal. These integrals, of functions with values in X or $[X]$, may be defined just as in classical analysis.

First let us consider some aspects of the situation when $T \in [X]$. In this case $\sigma(T)$ is bounded, and we have a power series formula for R_λ when $|\lambda| > r_\sigma(T)$; see (5.2–4). If C denotes a simple closed contour, oriented counterclockwise, enclosing the circle $|\lambda| = r_\sigma(T)$, we can integrate (5.2–4) term by term around the contour, and we obtain

$$I = \frac{1}{2\pi i}\int_C R_\lambda \, d\lambda, \tag{5.6–1}$$

because of the fact that the integral of each term of the series is 0, except for the first one. In a similar way,

$$T^p = \frac{1}{2\pi i}\int_C \lambda^p R_\lambda \, d\lambda \qquad p = 0, 1, 2, \cdots. \tag{5.6–2}$$

Since R_λ is analytic except at the points of $\sigma(T)$, it is clear that the integral in (5.6–2) is unchanged in value if we deform the contour C in any manner, so long as it continues to enclose $\sigma(T)$. We may even replace the single contour C by several nonintersecting closed contours, provided that no one of them is inside any other and $\sigma(T)$ lies in the union of their interiors.

We now propose to define a certain class of complex-valued analytic functions and to associate with each such function f an element of the operator space $[X]$. The operator associated with f will be denoted by $f(T)$. The feature of greatest importance in this association is that the correspondence between f and $f(T)$ preserves the basic algebraic operations. That is, the operators corresponding to $f + g$, αf, and fg, respectively, are $f(T) + g(T)$, $\alpha f(T)$, and $f(T)g(T)$. In particular, then, since $fg = gf$, the operators $f(T)$ and $g(T)$ commute.

In what follows we use $\Delta(f)$ to denote the domain of definition of f.

Definition. Suppose $T \in [X]$. Let $\mathfrak{A}(T)$ be the class of complex-valued functions f such that: (1) $\Delta(f)$ is an open set in the complex plane, and it contains $\sigma(T)$; (2) f is differentiable at each point of $\Delta(f)$.

We do not insist that $\Delta(f)$ be connected. If $\Delta(f)$ has two components, for example, it is legitimate to have $f(\lambda) = \lambda^2$ on one component and $f(\lambda) = 0$ on the other. The definition means that f is analytic, in the sense of classical function theory, on each component of $\Delta(f)$, but there is no necessary connection, by analytic continuation or otherwise, between the values of f in different components. We say that f is *locally analytic* on $\sigma(T)$.

Next we introduce some terminology and notation in connection with contour integrals.

Definition. A set D in the complex plane is called a *Cauchy domain* if: (1) it is open; (2) it has a finite number of components, the closures of any two of which are disjoint; (3) the boundary of D is composed of a finite positive number of closed rectifiable Jordan curves, no two of which intersect.

If C is one of the curves forming part of the boundary of D, the positive orientation of C is clockwise or counterclockwise according to whether the points of D near a point of C are outside or inside of C. The positively oriented boundary of D is denoted by $+ B(D)$; with the reverse orientation it is denoted by $- B(D)$.

For our purposes the following are typical examples of bounded Cauchy domains. Unbounded Cauchy domains will be used later.

Example 1. The set $\{\lambda : |\lambda| < 1\} \cup \{\lambda : 2 < |\lambda| < 3\}$.

Example 2. The set $\{\lambda : |\lambda| < 1\} \cup \{\lambda : |\lambda - 3| < 1\} \cup \{\lambda : |\lambda - 6| < 1\}$.

We are now ready to define an operator $f(T)$ corresponding to $f \in \mathfrak{A}(T)$. The definition is

$$f(T) = \frac{1}{2\pi i} \int_{+B(D)} f(\lambda) R_\lambda \, d\lambda, \tag{5.6–3}$$

where D is any bounded Cauchy domain such that $\sigma(T) \subset D$ and $\bar{D} \subset \Delta(f)$. Concerning this definition several comments are needed. (1) When f is given, there exists a Cauchy domain of the required sort. This is intuitively plausible; a detailed proof may be based on the construction given in Taylor, **2** (Theorem 3.3 and proof, page 65). (2) The integral in (5.6–3) has a value independent of the particular choice of D. This is so by an application of Cauchy's theorem. For, if D_1 and D_2 are two Cauchy domains of the sort considered, we have $\sigma(T) \subset D_1 \cap D_2$, and there exists a bounded Cauchy domain D such that $\sigma(T) \subset D$ and $\bar{D} \subset D_1 \cap D_2$. Now $D_1 - \bar{D}$ is a bounded Cauchy domain, and its oriented boundary consists of $+ B(D_1)$ and $- B(D)$. Moreover, $f(\lambda)R_\lambda$ has no singularities in $D_1 - \bar{D}$ or on its boundary, and hence, by Cauchy's theorem, the integral of $f(\lambda)R_\lambda$ over $+ B(D_1)$ is equal to the integral over $+ B(D)$. The same result holds with D_2 in place of D_1, and so our assertion is justified. (3) The integral in (5.6–3) is unchanged in value if we replace f by any other member of $\mathfrak{A}(T)$, say g, such that $f(\lambda) = g(\lambda)$ at each point λ of an open set containing $\sigma(T)$. This follows from the foregoing remarks.

We now show that the operator corresponding to $f_1 f_2$ is $f_1(T) f_2(T)$. The proof depends heavily on the equation $R_\lambda - R_\mu = (\mu - \lambda) R_\lambda R_\mu$ (see Theorem 5.1–C). Choose bounded Cauchy domains D_1, D_2 such that $\sigma(T) \subset D_1$, $\bar{D}_1 \subset D_2$, $\bar{D}_2 \subset \Delta(f_1) \cap \Delta(f_2)$. Then we express $f_1(T)$ as an integral with respect to λ over $+ B(D_1)$ and $f_2(T)$ as an integral with respect to μ over $+ B(D_2)$. Then we can write

$$f_1(T) f_2(T) = \frac{1}{2\pi i} \int_{+B(D_1)} f_1(\lambda) \left\{ \frac{1}{2\pi i} \int_{+B(D_2)} f_2(\mu) R_\lambda R_\mu \, d\mu \right\} d\lambda. \tag{5.6–4}$$

In this expression we replace $R_\lambda R_\mu$ by

$$\frac{R_\lambda - R_\mu}{\mu - \lambda} = \frac{R_\lambda}{\mu - \lambda} + \frac{R_\mu}{\lambda - \mu},$$

and invert the order of the iterated integration where the second fraction is concerned. Since $\lambda \in D_2$ and μ is not in $\bar{D}_1$, we have

$$\frac{1}{2\pi i} \int_{+B(D_2)} \frac{f_2(\mu)}{\mu - \lambda} \, d\lambda = f_2(\lambda), \qquad \frac{1}{2\pi i} \int_{+B(D_1)} \frac{f_1(\lambda)}{\lambda - \mu} \, d\lambda = 0.$$

Hence we obtain from (5.6–4) the desired result

$$f_1(T)f_2(T) = \frac{1}{2\pi i}\int_{+B(D_1)} f_1(\lambda)f_2(\lambda)R_\lambda\, d\lambda = (f_1f_2)(T).$$

It is obvious from (5.6–3) that $(f+g)(T) = f(T) + g(T)$ and $(\alpha f)(T) = \alpha f(T)$.

It is instructive to regard the association of $f(T)$ with f as an algebraic homomorphism. But in order to be precise about this it is first of all necessary to convert $\mathfrak{A}(T)$ into an algebra by an appropriate equivalence relation. We define two functions f, g as being equivalent (relative to T) if $f(\lambda) = g(\lambda)$ on some open set containing $\sigma(T)$. Then $\mathfrak{A}(T)$ is divided into equivalence classes, and these classes form a commutative algebra with a unit element if we define the algebraic operations in an obvious way, using representative functions. We have already noted that $f(T)$ is unchanged if f is replaced by an equivalent function.

Theorem 5.6–A. *The mapping $f \to f(T)$ by the formula* (5.6–3) *is an algebraic homomorphism of the algebra of the equivalence classes of $\mathfrak{A}(T)$ into the algebra $[X]$. This mapping carries the function $f(\lambda) \equiv 1$ into I and the function $f(\lambda) \equiv \lambda$ into T.*

The proof is covered by the preceding discussion. The last two assertions are justified by (5.6–1) and 5.6–2).

The reader may have noticed that the formula defining $f(T)$ has a striking appearance when for heuristic effect we write R_λ in the form $\dfrac{1}{\lambda - T}$:

$$f(T) = \frac{1}{2\pi i}\int_{+B(D)} \frac{f(\lambda)}{\lambda - T}\, d\lambda.$$

In formal structure this is just Cauchy's formula, with T in place of a complex number. We refer to the use of the homomorphism $f \to f(T)$ and the consequences flowing out of it as an operational calculus for T.

One use of the operational calculus is that it enables us to compute inverse operators in certain situations.

Theorem 5.6–B. *Suppose $T \in [X]$ and $f \in \mathfrak{A}(T)$. Suppose $f(\lambda) \neq 0$ when $\lambda \in \sigma(T)$. Then $f(T)$ sets up a one-to-one mapping of X onto all of X, with inverse $g(T)$, where g is any member of $\mathfrak{A}(T)$ equivalent to the reciprocal of $f(\lambda)$.*

PROOF. From $f(\lambda)g(\lambda) \equiv 1$ on a neighborhood of $\sigma(T)$ we infer $f(T)g(T) = g(T)f(T) = I$, and the conclusion follows.

We now give an example which illustrates the operational calculus for a particular operator.

***Example* 3.** Let $X = C[0, 1]$, and consider the very special Volterra-type operator T, where $Tx = y$ means

$$y(s) = \int_0^s x(t)\, dt.$$

It is easily verified by induction that $T^{n+1}x = y$ means

$$y(s) = \frac{1}{n!}\int_0^s (s - t)^n x(t)\, dt.$$

In this case the series (5.2–4) for R_λ converges whenever $\lambda \neq 0$ (i.e., $\sigma(T)$ is the single point $\lambda = 0$). By using this series we find that $x = R_\lambda y$ means

$$x(s) = \frac{1}{\lambda} y(s) + \frac{1}{\lambda^2}\int_0^s e^{(s-t)/\lambda} y(t)\, dt. \tag{5.6–5}$$

In this case $\mathfrak{A}(T)$ consists of functions analytic in a neighborhood of $\lambda = 0$. If f is such a function, we can deduce the meaning of $f(T)$ from (5.6–3) and (5.6–5). The relation $y = f(T)x$ can be written

$$y(s) = \frac{1}{2\pi i}\oint f(\lambda)\left\{\frac{1}{\lambda} x(s) + \frac{1}{\lambda^2}\int_0^s e^{(s-t)/\lambda} x(t)\, dt\right\} d\lambda,$$

or

$$y(s) = f(0)x(s) + \int_0^s x(t)\left\{\frac{1}{2\pi i}\oint \frac{f(\lambda)}{\lambda^2} e^{(s-t)/\lambda}\, d\lambda\right\} dt, \tag{5.6–6}$$

where $\oint$ denotes integration counterclockwise around some sufficiently small circle $|\lambda| = r$. We also write this in the form

$$y(s) = f(0)x(s) + \int_0^s F(s - t)x(t)\, dt, \tag{5.6–7}$$

where

$$F(u) = \frac{1}{2\pi i}\oint \frac{f(\lambda)}{\lambda^2} e^{u/\lambda}\, d\lambda. \tag{5.6–8}$$

The function F turns out to be an entire function of exponential type. If $f(\lambda)$ is expressed as a power series in λ, with radius of convergence α, the type of F is exactly $1/\alpha$. See problem 1 for more on this subject.

We can use the operational calculus for this particular operator to solve the differential equation

$$y^{(n)}(s) + a_1 y^{(n-1)}(s) + \cdots + a_n y(s) = x(s)$$

with the initial conditions

$$y(0) = y'(0) = \cdots = y^{(n-1)}(0) = 0,$$

where $x \in C[0, 1]$ and the coefficients $a_1, \cdots, a_n$ are constants. It is easy to see that y is a solution of this differential equation and satisfies the initial condition if and only if

$$(I + a_1 T + \cdots + a_n T^n)y = T^n x.$$

Now let $g(\lambda) = 1 + a_1\lambda + \cdots + a_n\lambda^n$. The problem is to solve $g(T)y = T^n x$ for y. Since $g(0) \neq 0$, we know by Theorem 5.6–B that $g(T)$ has an inverse. The operational calculus shows that $y = f(T)x$, where $f(\lambda) = \lambda^n/g(\lambda)$. Hence, by (5.6–6), the solution of the problem is

$$y(s) = \int_0^s x(t) \left\{\frac{1}{2\pi i} \oint \frac{\lambda^{n-2}}{g(\lambda)} e^{(s-t)/\lambda}\, d\lambda\right\} dt,$$

where the contour encloses $\lambda = 0$, and all the zeros of $g(\lambda)$ are outside of the contour. By the change of variable $z = \lambda^{-1}$ this solution can be put in the form

$$y(s) = \int_0^s x(t) \left\{\frac{1}{2\pi i} \oint \frac{e^{(s-t)z}}{z^n + a_1 z^{n-1} + \cdots + a_n}\, dz\right\} dt,$$

where the contour is counterclockwise and encloses all the zeros of the polynomial in the denominator. The integral may be evaluated by computing residues.

It is desirable to develop a generalization of the formula (5.6–3) so as to yield an operational calculus for T when T is any closed linear operator with domain and range in X. It turns out that such a development is possible provided that the resolvent set $\rho(T)$ is not empty. When we give up the condition that $T \in [X]$, the spectrum of T need no longer be compact; it might be empty or it might be the whole plane. However, we assume explicitly that $\sigma(T)$ is *not* the whole plane; we also assume that T is closed. This permits us to utilize Theorem 5.1–C.

Definition. By $\mathfrak{A}_\infty(T)$ we mean the class of complex-valued functions f such that: (1) $\Delta(f)$ is an open set in the complex plane which contains $\sigma(T)$ and is such that the complement of $\Delta(f)$ is compact; (2) f is differentiable in $\Delta(f)$ and $f(\lambda)$ is bounded as $|\lambda| \to \infty$.

We know from function theory that $f(\lambda)$ approaches a finite limit as $|\lambda| \to \infty$, and we denote this limit by $f(\infty)$. We may then say that f is locally analytic on $\sigma(T)$ and at ∞.

We may define an equivalence relation in $\mathfrak{A}_\infty(T)$: two functions are equivalent if they agree on a neighborhood of $\sigma(T)$ and also on a neighborhood of ∞. As before, the equivalence classes form a commutative algebra in an obvious way. The function $f(\lambda) \equiv 1$ determines a unit for the algebra.

In seeking the proper replacement for (5.6–3) we observe that Cauchy's formula for an element of $\mathfrak{A}_\infty(T)$ holds in the form

$$f(\xi) = f(\infty) + \frac{1}{2\pi i}\int_{+B(D)} \frac{f(\lambda)}{\lambda - \xi}\,d\lambda, \tag{5.6–9}$$

where D is an *unbounded* Cauchy domain such that $\bar{D} \subset \Delta(f)$ and $\xi \in D$. Note that the complement of an unbounded Cauchy domain is compact; in fact, there is just one unbounded component of such a domain, and its complement is compact.

The appropriate definition of $f(T)$ when $f \in \mathfrak{A}_\infty(T)$ is

$$f(T) = f(\infty)I + \frac{1}{2\pi i}\int_{+B(D)} f(\lambda)R_\lambda\,d\lambda, \tag{5.6–10}$$

where D is an unbounded Cauchy domain such that $\sigma(T) \subset D$ and $\bar{D} \subset \Delta(f)$. The three comments made after (5.6–3) have counterparts in the present situation. Observe that $f(T) \in [X]$ even though T need not be in $[X]$.

As an illustration of $\Delta(f)$ and D, suppose $\sigma(T)$ is the entire real axis. Then, for some positive r and ϵ, $\Delta(f)$ must include all points for which $|\lambda| > r$ and all points for which the imaginary part of λ is in absolute value less than ϵ. We could then take D to be the union of the sets $\{\lambda : |\lambda| > 2r\}$, $\{\lambda : |\lambda| \leqslant 2r \text{ and } Im\,|\lambda| < \epsilon/2\}$.

As in the case of Theorem 5.6–A, the mapping $f \to f(T)$ defined by (5.6–10) is an algebraic homomorphism of the algebra of equivalence classes of $\mathfrak{A}_\infty(T)$ into the algebra $[X]$, and the algebra preserves the unit element, i.e., $f(\lambda) \equiv 1$ maps into $f(T) = I$. To see the truth of this last assertion let D_1 be the complement of the D in (5.6–10). Then D_1 is a bounded Cauchy domain, $\bar{D}_1 \subset \rho(T)$, and $+ B(D) = - B(D_1)$. Since R_λ is analytic on $\rho(T)$, the integral in (5.6–10) vanishes if $f(\lambda) \equiv 1$. In this case $f(\infty) = 1$, and we get $f(T) = I$. The proof that $(f_1f_2)(T) = {}_1(T)f_2(T)$ is similar to the corresponding proof based on (5.6–3), and we leave the argument to the reader. One must use (5.6–9).

If T is not in $[X]$, we must use (5.6–10) instead of (5.6–3). But, if $T \in [X]$ and $f \in \mathfrak{A}_\infty(T)$, the operator $f(T)$ given by (5.6–10) is the same as that given by (5.6–3). To prove this, choose for the D in (5.6–10) the union of a bounded Cauchy domain D_1 and the exterior of a very large circle C which encloses D_1, where $\sigma(T) \subset D_1$ and $\bar{D}_1 \subset \Delta(f)$. Then the integral over $+ B(D)$ in (5.6–10) becomes the integral over $+ B(D_1)$ plus an integral around C. Since $\|R_\lambda\| \to 0$ as $|\lambda| \to \infty$ (when $T \in [X]$), it is easy to prove by standard methods that the integral around C cancels the term $f(\infty)I$.

Corresponding to Theorem 5.6–B we have:

Theorem 5.6–C. *If T is closed, $f \in \mathfrak{A}_\infty(T)$ and f has no zeros on $\sigma(T)$ or at ∞, the operator $f(T)$ has an inverse which belongs to $[X]$. This inverse is $g(T)$, where g is any member of $\mathfrak{A}_\infty(T)$ equivalent to the reciprocal of $f(\lambda)$.*

The proof is the same as for Theorem 5.6–B.

A polynomial of degree $n \geqslant 1$ does not belong to $\mathfrak{A}_\infty(T)$. Nevertheless, it is convenient to be able to deal with polynomials in connection with the operational calculus. We consider now how this is to be done.

If $P(\lambda)$ is a polynomial of degree n, we define $P(T)$ in the obvious way, by putting T^k in place of λ^k in the expression for $P(\lambda)$; the domain of $P(T)$ is $\mathscr{D}_n(T)$, as defined at the beginning of § 5.41. We have several lemmas which are useful in dealing with polynomials.

Lemma 5.6–D. *Suppose $f \in \mathfrak{A}_\infty(T)$, and suppose either (a) that f has a zero of order m at ∞, or (b) that f vanishes identically in a neighborhood of ∞. Let P be a polynomial of degree n, where $0 < n \leqslant m$ in case a and $0 < n$ in case b. Let $F(\lambda) = P(\lambda)f(\lambda)$. Then $F \in \mathfrak{A}_\infty(T)$, the range of $f(T)$ lies in $\mathscr{D}_n(T)$, and $F(T) = P(T)f(T)$.*

For an indication of the proof see problem 6.

Lemma 5.6–E. *Suppose $f \in \mathfrak{A}_\infty(T)$, $\alpha \in \rho(T)$, and an integer $n \geqslant 0$ are given, and suppose that $g \in \mathfrak{A}_\infty(T)$, where $g(\lambda) = (\alpha - \lambda)^n f(\lambda)$. (Note that this hypothesis is essentially a condition on the behavior of f near $\lambda = \infty$.) Then $f(T)\mathscr{D}_k(T) \subset \mathscr{D}_{k+n}(T)$ when $k \geqslant 0$.*

PROOF. If $x \in \mathscr{D}_k(T)$ we can write $x = (R_\alpha)^k y$ for some y. Then $f(T)x = (R_\alpha)^n g(T)(R_\alpha)^k y = (R_\alpha)^{n+k} g(T) y \in \mathscr{D}_{n+k}(T)$.

Lemma 5.6–F. *If $f \in \mathfrak{A}_\infty(T)$ and $P(\lambda)$ is a polynomial of degree $n \geqslant 1$, then $f(T)P(T)x = P(T)f(T)x$ if $x \in \mathscr{D}_n(T)$.*

PROOF. Let $G(\lambda) = (\alpha - \lambda)^{-n}P(\lambda)$. Then $G(T) = P(T)(R_\alpha)^n$ (see problem 6). If $x \in \mathscr{D}_n(T)$, we can write $x = (R_\alpha)^n y$. We know $f(T)x \in \mathscr{D}_n(T)$ (Lemma 5.6–E). Then $P(T)f(T)x = P(T)f(T)(R_\alpha)^n y = G(T)f(T)y = f(T)P(T)(R_\alpha)^n y$, and the proof is complete.

We see next how to express $P(T)x$ as an integral.

Theorem 5.6–G. *Suppose $\alpha \in \rho(T)$. Let D be any unbounded Cauchy domain such that $\sigma(T) \subset D$ and α is not in $\bar{D}$. Then, if $P(\lambda)$ is a polynomial of degree n and $x \in \mathscr{D}_n(T)$,*

$$P(T)x = \frac{1}{2\pi i}\int_{+B(D)} \frac{P(\lambda)}{(\lambda - \alpha)^{n+1}}(T - \alpha)^{n+1} R_\lambda x \, d\lambda. \tag{5.6–11}$$

If $f \in \mathfrak{A}_\infty(T)$ *we can also compute* $f(T)$ *by this formula for any* $n \geqslant 0$, *by putting* $f(\lambda)$ *in place of* $P(\lambda)$ *and* $f(T)$ *in place of* $P(T)$, *provided we choose* D *so that* $\sigma(T) \subset D$ *and* $\bar{D} \subset \Delta(f)$.

PROOF. We start from the formula

$$(T - \alpha)^{n+1}R_\lambda x = (\lambda - \alpha)^{n+1}R_\lambda x - \sum_{k=0}^{n} (\lambda - \alpha)^{n-k}(T - \alpha)^k x,$$

which is easily established by induction. The evaluation of the integral in (5.6–11) then becomes a matter of evaluation of familiar integrals, from which the final results are easily obtained. We leave details to the reader.

The next theorem is a generalization of Theorem 5.6–C.

Theorem 5.6–H. *Suppose* $f \in \mathfrak{A}_\infty(T)$ *and* $f(\lambda) \neq 0$ *if* $\lambda \in \sigma(T)$, *but that* $f(\infty) = 0$, *the zero at* ∞ *being of finite order* m. *Then* $f(T)$ *has an inverse, the range of* $f(T)$ *is* $\mathscr{D}_m(T)$, *and for* $x \in \mathscr{D}_m(T)$ *we have*

$$[f(T)]^{-1}x = \frac{1}{2\pi i}\int_{+B(D)} \{f(\lambda)(\lambda - \alpha)^{m+1}\}^{-1}(T - \alpha)^{m+1}R_\lambda x\, d\lambda. \tag{5.6–12}$$

Here $\alpha \in \rho(T)$; α *and the unbounded Cauchy domain* D *are to be chosen so that* $\sigma(T) \subset D$, $\bar{D} \subset \Delta(f)$, α *is not in* $\bar{D}$, *and* $f(\lambda) \neq 0$ *if* $\lambda \in \bar{D}$.

PROOF. Let $g(\lambda) = (\alpha - \lambda)^m f(\lambda)$. Then g has no zeros on $\sigma(T)$ or at ∞, so that $g(T)$ has an inverse belonging to $[X]$. Now $f(T) = (R_\alpha)^m g(T)$. Hence $f(T)$ has the inverse $[g(T)]^{-1}(\alpha - T)^m$ with domain $\mathscr{D}_m(T)$. We calculate $[f(T)]^{-1}x = [g(T)]^{-1}(\alpha - T)^m x$ by using (5.6–11), putting $n = 0$, $1/g(\lambda)$ in place of $P(\lambda)$, and $(\alpha - T)^m x$ in place of x. The result is (5.6–12).

We now consider an example of a closed operator with unbounded spectrum.

Example 4. Let $X = C[0, 2\pi]$. Let $\mathscr{D}(T)$ be the set of continuously differentiable functions $x \in X$ such that $x(0) = x(2\pi)$, and let $Tx = y$ mean $y(s) = -ix'(s)$. To compute R_λ we solve the differential equation

$$x'(s) - i\lambda x(s) = -iy(s)$$

with the condition $x(0) = x(2\pi)$ on the solution. We find that the values $\lambda = 0, \pm 1, \pm 2, \cdots$ are in the point spectrum, the eigenfunctions corresponding to $\lambda = n$ being multiples of e^{ins}. All other values of s are in $\rho(T)$, with $x = R_\lambda y$ expressed by

$$x(s) = \frac{e^{i\pi\lambda}}{2\sin\pi\lambda}\int_0^{2\pi} e^{i\lambda(s-t)}y(t)\,dt - i\int_0^s e^{i\lambda(s-t)}y(t)\,dt. \tag{5.6–13}$$

Now suppose $f \in \mathfrak{A}_\infty(T)$. This means that f is analytic at ∞ and also in some neighborhood of each of the points of $\sigma(T)$. For the purpose of computing $f(T)$ we may assume that $f(\lambda) = \sum_0^\infty a_n \lambda^{-n}$ when $|\lambda| > N + {}^1/_3$, where N is some positive integer, and that near $\lambda = k$ $(k = 0, \pm 1, \cdots, \pm N)$ f is given by $f(\lambda) = g_k(\lambda)$, where g_k is analytic at $\lambda = k$. In computing $f(T)$ by formula (5.6–10) we can take D to consist of the exterior of the circle C: $|\lambda| = N + {}^1/_2$ and the union of the interiors of circles C_k: $|\lambda - k| < {}^1/_4$, $k = 0, \pm 1, \cdots, \pm N$. The contribution to $f(T)x$ from integration around C_k turns out to be

$$\frac{g_k(k)}{2\pi} \int_0^{2\pi} x(t) e^{ik(s-t)}\, dt = g_k(k)\xi_k e^{iks},$$

where ξ_k is the Fourier coefficient of $x(t)$ with respect to e^{ikt}. The general formula for $y = f(T)x$ is a bit lengthy, and we forego writing it out here. However, in the special case where $f(\lambda) \equiv a_0$ when $|\lambda| > N + {}^1/_3$, things are much simpler, and $y = f(T)x$ is expressed by

$$y(s) = a_0 x(s) + \sum_{k=-N}^{N} (b_k - a_0)\xi_k e^{iks}, \tag{5.6–14}$$

where $b_k = g_k(k)$. The coefficients a_0 and b_k can be assigned arbitrarily. From Theorem 5.6–C we get the interesting result that, if a_0 and the b_k's are all different from zero, the solution of (5.6–14) for x is

$$x(s) = \frac{1}{a_0} y(s) + \sum_{k=-N}^{N} \left(\frac{1}{b_k} - \frac{1}{a_0}\right)\eta_k e^{iks},$$

where the η_k's are the Fourier coefficients of y.

If we impose special conditions on T, there may be various other ways besides that indicated by (5.6–10) for developing an operational calculus. The cases in which T is such that $\sigma(T)$ lies in a half plane (Hille, **1**, Chapter 15) or in a strip (Bade, **1**) are of great interest, and in these cases very interesting and useful operational calculi have been developed.

PROBLEMS

1. If $f(\lambda) = \sum_0^\infty a_n \lambda^n$ in (5.6–6), show that $F(u) = \sum_0^\infty a_{n+1}(u^n/n!)$ in (5.6–8). If the radius of convergence of the f series is α, F is of exponential type $1/\alpha$. See R. P. Boas, **1**, page 839. We see that f determines F; conversely, F

determines f, except that a_0 is left arbitrary. Any Volterra-type integral operator with kernel $F(s - t)$, where F is an entire function of exponential type, is an operator $f(T)$, where f is analytic at $\lambda = 0$ and $f'(0) = 0$.

2. If $f(\lambda) = \sum_0^\infty a_n\lambda^n$ in (5.6–6), show that $f(T) = \sum_0^\infty a_nT^n$, the series converging in $[X]$.

3. Let $X = C[a, b]$ (a finite interval). Let T be the differentiation operator, $Tx = x'$, where $\mathscr{D}(T)$ is the set of those $x \in X$ such that $x(a) = 0$ and the derivative x' also belongs to X. This operator is closed, and $\sigma(T)$ is empty. Show that $x = R_\lambda y$ means

$$x(s) = -\int_a^s e^{\lambda(s-t)}y(t)\,dt.$$

If $f \in \mathfrak{A}_\infty(T)$ is defined by $f(\lambda) = \sum_0^\infty a_n\lambda^{-n}$, show that $f(T)x = y$ means

$$y(s) = a_0x(s) + \int_a^s F(s - t)x(t)\,dt,$$

where $F(u) = \sum_0^\infty a_{n+1}(u^n/n!)$. The situation here is closely related to that in Example 3, because $y = T^{-1}x$ means $y(s) = \int_a^s x(t)\,dt$.

4. When T is closed and $\rho(T)$ is not empty, $P(T)$ is closed. Outline of proof: If $\alpha \in \rho(T)$, let $A = (T - \alpha)^{-1}$. Write $P(\lambda) = \sum_0^n b_k(\lambda - \alpha)^{n-k}$, $p(\mu) = \sum_0^n b_k\mu^k$, $b_0 \neq 0$. Then $P(T)x = p(A)(T - \alpha)^nx$ if $x \in \mathscr{D}_n(T)$. Now $(T - \alpha)^n$ is closed, for it is the inverse of A^n. It can then be proved that $P(T)$ is closed. See Taylor, **2**, Theorem 6.1. If $\rho(T)$ is void, it can occur that T is closed but T^2 is not. Here is an example: Let $X = l^2 \times l^2$. If $(x, y) \in X$ let $\|(x, y)\|^2 = \|x\|^2 + \|y\|^2$. If $x = \{\xi_k\}$ and $y = \{\eta_k\}$, define $T(x, y) = (\{k^{-2}\eta_k\}, \{k\xi_k\})$, with $\mathscr{D}(T)$ consisting of all (x, y) with $y \in l^2$ and $\{k\xi_k\} \in l^2$.

5. Suppose $\alpha \in \rho(T)$. Then the function $f(\lambda) = (\alpha - \lambda)^{-n}(n \geqslant 1)$ belongs to $\mathfrak{A}_\infty(T)$, and $f(T) = (R_\alpha)^n$. Prove this for $n = 1$ directly from (5.6–10), using the fact that $R_\lambda = R_\alpha + (\alpha - \lambda)R_\lambda R_\alpha$. For general n the result can be proved by induction.

6. To prove Lemma 5.6–D let $g(\lambda) = (\alpha - \lambda)^nf(\lambda)$, where $\alpha \in \rho(T)$. Then $f(T) = (R_\alpha)^ng(T)$ (use problem 5). Let $G(\lambda) = (\alpha - \lambda)^{-n}P(\lambda)$ and show that $G(T) = P(T)(R_\alpha)^n$ by expressing $P(\lambda)$ as a sum of powers of $\alpha - \lambda$. The relation $F(T) = P(T)f(T)$ now follows at once.

7. Let $P(\lambda)$ by a polynomial of degree $n \geqslant 1$ all of whose zeros lie in $\rho(T)$. Then $P(T)$ has X for its range, and it has an inverse belonging to $[X]$ and given by

$$[P(T)]^{-1} = \frac{1}{2\pi i}\int \frac{1}{P(\lambda)}R_\lambda\,d\lambda,$$

the integration being extended clockwise around a set of nonoverlapping circles, one centered at each zero of $P(\lambda)$, each circle and its interior lying in $\rho(T)$. Method of proof: Let $f(\lambda) = 1/P(\lambda)$ and apply Theorem 5.6–H. A comparison of (5.6–12) for this case with (5.6–11) shows that $[f(T)]^{-1}x = P(T)x$, whence $[P(T)]^{-1} = f(T)$. Then apply (5.6–10).

8. Let T be the operator of problem 3. Use the formula in problem 7 to solve $P(T)y = x$ for y, where $P(\lambda)$ is a polynomial of degree n. Compare with the last part of Example 3.

9. Let $X = H^2$ (see Example 8, § 3.11). Define $Tx(t) = tx(t)$ ($x \in X$, t the complex variable). Then $T \in [X]$, $\sigma(T) = \{\lambda : |\lambda| \leqslant 1\}$. If $f \in \mathfrak{A}(T)$, $f(T)x(t) = f(t)x(t)$.

10. Let $X = L(0, \infty)$, and let T be the differentiation operator $Tx = x'$, with $\mathscr{D}(T)$ determined by those functions in $\mathscr{L}(0, \infty)$ which are absolutely continuous on $[0, a]$ for every finite $a > 0$ and whose derivative is again in $\mathscr{L}(0, \infty)$. It can be seen from Example 4 in § 1.5 that $\sigma(T)$ consists of all λ for which the real part of λ is $\leqslant 0$. Also, $x = R_\lambda y$ is expressed by

$$x(s) = \int_0^\infty y(t)e^{\lambda(s-t)}\,dt.$$

Discuss the nature of $\mathfrak{A}_\infty(T)$ and the form of $f(T)$ in this case.

5.7 Spectral Sets and Projections

The results of this section are related to the concept of reducibility (see § 5.4). As we shall see, we can obtain a pair of subspaces which reduce T completely if we can split the spectrum of T into two parts of a suitable nature. We assume throughout the section that X is a complex Banach space and that T is a closed linear operator with domain and range in X. We also assume as always that $\rho(T)$ is not empty.

In addition to the concept of the spectrum of T, we need the concept of the *extended spectrum* of T. This is a point set in the *extended* complex plane (i.e., the one-point compactification of the ordinary plane, by adjunction of the point ∞). We denote the extended spectrum of T by $\sigma_e(T)$. It is defined to be the same as $\sigma(T)$ if $T \in [X]$ and to consist of $\sigma(T)$ and the point ∞ if T is not in $[X]$. Observe that $\sigma_e(T)$ is always closed and nonempty.

Definition. A subset σ of $\sigma_e(T)$ is called a *spectral set* of T if it is both open and closed in the relative topology of $\sigma_e(T)$ as a subset of the extended plane. This is the same as requiring that both σ and $\sigma_e(T) - \sigma$ be closed in the extended plane.

An isolated point of $\sigma_e(T)$ is of course a spectral set. If σ is a spectral set and if one of the sets σ, $\sigma_e(T) - \sigma$ contains ∞, then the other one is bounded as a subset of the ordinary plane.

If σ is a spectral set of T, we can define an $f \in \mathfrak{A}_\infty(T)$ such that $f(\lambda) = 1$ on a neighborhood of σ while $f(\lambda) = 0$ on a neighborhood of $\sigma_e(T) - \sigma$. We then denote the operator $f(T)$ by E_σ. When σ_1 occurs as a subscript on E or X in what follows, we write $\sigma(1)$ instead of σ_1, for convenience in printing. Likewise with σ_2. Since $f(\lambda)f(\lambda) = f(\lambda)$, the homomorphism shows that $E_\sigma E_\sigma = E_\sigma$, so that E_σ is a projection. We call it *the projection associated with* σ. Evidently the function $1 - f(\lambda)$ is related to $\sigma_e(T) - \sigma$ in the way that f is related to σ; consequently $I - E_\sigma$ is the projection associated with the spectral set $\sigma_e(T) - \sigma$. If we denote σ and $\sigma_e(T) - \sigma$ by σ_1 and σ_2 respectively, we see that $E_{\sigma(1)} + E_{\sigma(2)} = I$, and $E_{\sigma(1)}E_{\sigma(2)} = E_{\sigma(2)}E_{\sigma(1)} = 0$. Let $X_{\sigma(i)}$ be the range of $E_{\sigma(i)}$; it is a closed subspace. We see that $X_{\sigma(1)} \cap X_{\sigma(2)} = (0)$ and $X_{\sigma(1)} \oplus X_{\sigma(2)} = X$.

To show that T is completely reduced by $(X_{\sigma(1)}, X_{\sigma(2)})$ it suffices to show that $E_{\sigma(1)}\mathscr{D}(T) \subset \mathscr{D}(T)$ and $E_{\sigma(1)}Tx = TE_{\sigma(1)}x$ if $x \in \mathscr{D}(T)$ (see § 5.4). These things follow from Lemmas 5.6–E (with $n = 0$, $k = 1$) and 5.6–F.

The foregoing considerations can easily be generalized to the extent stated in the following theorem, proof of which is left to the reader.

Theorem 5.7–A. *Suppose* $\sigma_e(T) = \sigma_1 \cup \cdots \cup \sigma_n$, *where* $\sigma_1, \cdots, \sigma_n$ *are pairwise disjoint spectral sets of* T. *Let* $E_{\sigma(i)}$ *be the projection associated with* σ_i *and let* $X_{\sigma(i)}$ *be the range of* $E_{\sigma(i)}$. *Then* T *is completely reduced by* $(X_{\sigma(1)}, \cdots, X_{\sigma(n)})$, *i.e.*, $I = E_{\sigma(1)} + \cdots + E_{\sigma(n)}$, $E_{\sigma(i)}E_{\sigma(j)} = 0$ *if* $i \neq j$, $\mathscr{D}(T)$ *is invariant under* $E_{\sigma(i)}$, *and* $X_{\sigma(i)}$ *is invariant under* T.

In the next theorem we examine the restriction of T to the range of E_σ.

Theorem 5.7–B. *Let* σ *be a spectral set of* T, *and let* T_1 *be the restriction of* T *to the range* X_σ *of* E_σ (with $\mathscr{D}(T_1) = \mathscr{D}(T) \cap X_\sigma$). *Then*

a. $\sigma = \sigma_e(T_1)$.
b. $\sigma \cap P\sigma(T) = P\sigma(T_1)$.
c. $\sigma \cap C\sigma(T) = C\sigma(T_1)$.
d. $\sigma \cap R\sigma(T) = R\sigma(T_1)$.
e. If σ *does not contain* ∞, *then* $X_\sigma \subset \mathscr{D}_n(T)$ *for each* $n \geqslant 1$, *and* T_1 *is continuous on* X_σ.

PROOF. It is convenient to prove (*e*) first. When σ does not contain ∞, we have $E_\sigma = f(T)$, where one property of f is that $f(\lambda) = 0$ on a neighborhood of ∞. We can then apply Lemma 5.6–E, with $k = 0$ and n arbitrary, to conclude that $X_\sigma \subset \mathscr{D}_n(T)$. But now T_1 is defined on all of X_σ, and hence it is continuous, for T_1 is closed (since T is) and X_σ is complete. Here we use Theorem 4.2–I.

Next we prove (*a*). Here it is convenient to write $\sigma_1 = \sigma$, $\sigma_2 = \sigma_e(T) - \sigma_1$ and to denote by T_2 the restriction of T to $X_{\sigma(2)}$. Select any finite point μ not in σ_1. If f is the function used in obtaining $f(T) = E_{\sigma(1)}$, let $g(\lambda) = (\mu - \lambda)^{-1}f(\lambda)$. Then $g \in \mathfrak{A}_\infty(T)$, $fg = g$, and so $(\mu - T)g(T) = E_{\sigma(1)}$ and $E_{\sigma(1)}g(T) = g(T)$, from which it appears that $\mu - T$ and $g(T)$ are inverse to each other when restricted to $X_{\sigma(1)}$. Consequently $\mu \in \rho(T_1)$. This shows that every point of $\sigma(T_1)$ is a finite point of σ_1; likewise for $\sigma(T_2)$ and σ_2. Hence $\sigma(T_1) \cap \sigma(T_2)$ is empty. Now $\sigma(T) = \sigma(T_1) \cup \sigma(T_2)$ (by Theorem 5.4–C*a*). It therefore follows that $\sigma(T_1)$ is identical with the set of finite points of σ_1; likewise for $\sigma(T_2)$ and σ_2. In view of (*e*) this proves (*a*) for the case in which ∞ is not in σ_1. If ∞ *is* in σ_1, then T is not in $[X]$. Also, ∞ is not in σ_2, so T_2 is continuous on $X_{\sigma(2)}$, by *e*. It follows that T_1 cannot belong to $[X_{\sigma(1)}]$, for otherwise T would be in $[X]$. Therefore $\infty \in \sigma_e(T_1)$, and (*a*) is completely proved. Assertions *b*, *c* and *d* now follow from Theorem 5.4–C.

The following simple example is instructive.

Example 1. Let $X = l^1$ (l^2 would do just as well). Let $\{\lambda_n\}$ be a sequence of distinct nonzero numbers such that $\lambda_n \to 0$, and define Tx by $T\{\xi_k\} = \{\lambda_k\xi_k\}$. Then $\sigma(T)$ consists of the points $\lambda_1, \lambda_2, \cdots$ and 0. The resolvent is defined by $R_\lambda\{\xi_k\} = \{(\lambda - \lambda_k)^{-1}\xi_k\}$. Let E_k be the projection associated with the spectral set formed by the single point λ_k, and let Q_N be the projection associated with the spectral set consisting of λ_{N+1}, $\lambda_{N+2}, \cdots$ and 0. Let $\{u_n\}$ be the standard countable basis for l^1, i.e., $u_1 = (1, 0, 0, \cdots)$, $u_2 = (0, 1, 0, \cdots)$ etc. Then we have

$$x = \sum_1^\infty \xi_k u_k, \qquad E_k x = \xi_k u_k, \qquad Q_N x = \sum_{N+1}^\infty \xi_k u_k.$$

Note that $I = E_1 + \cdots + E_N + Q_N$. Also, $TE_k = \lambda_k E_k$, and so $T = \lambda_1 E_1 + \cdots + \lambda_N E_N + TQ_N$. It is interesting to see what happens as $N \to \infty$. We have $\|Q_N x\| \to 0$ for each x, and so, for each x,

$$x = \sum_1^\infty E_k x \qquad \text{(convergence in } X).$$

But $\|Q_N\| = 1$, and so we *cannot* write $I = \sum_1^\infty E_k$ (with convergence in $[X]$). However, $\|TQ_N\| \to 0$, and so we *do* have

$$T = \sum_1^\infty \lambda_k E_k \qquad \text{(convergence in } [X]).$$

We also have

$$R_\lambda x = \sum_1^\infty \frac{E_k x}{\lambda - \lambda_k}, \qquad \text{but not } R_\lambda = \sum_1^\infty \frac{E_k}{\lambda - \lambda_k}.$$

For another instructive example we revert to § 5.6.

Example 2. Consider the operator T of Example 4, § 5.6, where the space X is $C[0, 2\pi]$. For this operator $\sigma_e(T)$ consists of the points $0, \pm 1, \pm 2, \cdots$ and ∞. If E_n is the projection associated with the spectral set consisting of the single point $\lambda = n$, we easily find from (5.6–13) that $E_n x(s) = \xi_n e^{ins}$, where ξ_n is the Fourier coefficient of $x(t)$ with respect to e^{int}. If Q_N is the projection associated with the spectral set consisting of $\pm (N + 1), \pm (N + 2), \cdots$ and ∞, we find that

$$Q_N x(s) = x(s) - \sum_{k=-N}^{N} \xi_k e^{iks}.$$

In this case it is not *always* true that $Q_N x \to 0$ as $N \to \infty$. It *is* true, however, if $x \in \mathscr{D}(T)$, by a standard theorem on Fourier series. Thus we can write

$$x = \sum_{-\infty}^{\infty} E_n x \tag{5.7–1}$$

when $x \in \mathscr{D}(T)$. We have $TE_n = nE_n$. In general it is not true that

$$Tx = \sum_{-\infty}^{\infty} nE_n x \tag{5.7–2}$$

when $x \in \mathscr{D}(T)$. But this is true if $x \in \mathscr{D}_2(T)$, since $E_n Tx = TE_n x = nE_n x$, and in this case we can use (5.7–1) with Tx in place of x.

It is interesting to see what happens in the foregoing example if we change the basic space X in which the operator T works. Suppose we take $X = L^2(0, 2\pi)$ and define $Tx = -ix'$, with $\mathscr{D}(T)$ the subspace of X determined by those x such that $x(s)$ is absolutely continuous on $[0, 2\pi]$, with $x(0) = x(2\pi)$ and such that $x'(s)$ belongs to $\mathscr{L}^2(0, 2\pi)$. The spectrum and the formula for the resolvent are just as before, and the formulas for E_n and Q_N are unchanged. But now (5.7–1) holds for all $x \in L^2(0, 2\pi)$. For, if we let $u_n(s) = e^{ins}/\sqrt{2\pi}$, the set $\{u_n\}(n = 0, \pm 1, \pm 2, \cdots)$ is a complete orthonormal set (see § 3.22), and (5.7–1) is just the standard expansion of x with respect to this orthonormal set. That is, $E_n x = (x, u_n)u_n$. Formula (5.7–2) is now true for every x in $\mathscr{D}(T)$, for $(Tx, u_n)u_n = (x, Tu_n)u_n = (x, nu_n)u_n = nE_n x$.

We return again to the general theory. It is easily seen that, if σ is a spectral set of T, then $E_\sigma = 0$ if and only if σ is empty, and $E_\sigma = I$ if and only if $\sigma = \sigma_e(T)$. Now, if $\mathscr{D}(T) \neq X$ and if $\sigma(T)$ is nonempty and bounded, then $\sigma(T)$ and (∞) are complementary nonempty spectral sets.

In this case, therefore, the projection associated with $\sigma(T)$ is neither 0 nor I.

We conclude this section with a theorem which is in a way a companion to Theorem 5.6–H.

Theorem 5.7–C. *Suppose that $\sigma(T)$ is bounded and that f is an element of $\mathfrak{A}_\infty(T)$ which vanishes on a neighborhood of ∞ but has no zeros on $\sigma(T)$. Then $f(T)$ has the same range and null manifold as the projection associated with $\sigma(T)$. In particular, $f(T)$ has no inverse if $\mathscr{D}(T) \neq X$.*

PROOF. We consider two disjoint parts of $\Delta(f)$: the component containing ∞, in which $f(\lambda) \equiv 0$, and the rest of $\Delta(f)$, which is a neighborhood of $\sigma(T)$. We define two functions g and h: $g(\lambda) = 0$ and $h(\lambda) = 1$ on the first part of $\Delta(f)$, $g(\lambda) = 1$ and $h(\lambda) = f(\lambda)$ on the other part. Note that h has no zeros on $\sigma(T)$ or at ∞, so that $h(T)$ has an inverse belonging to $[X]$. Note also that $g(T) = E_\sigma$, where $\sigma = \sigma(T)$. Now $fg = f$ and $gh = f$. Hence $f(T)E_\sigma = E_\sigma f(T) = f(T)$ and $E_\sigma h(T) = h(T)E_\sigma = f(T)$. The truth of the theorem follows from these relations, in view of the fact about the inverse of $h(T)$.

5.71 The Spectral Mapping Theorem

The following generalization of Theorem 5.2–D will be needed in § 5.8. We assume T closed.

Theorem 5.71–A. *If $f \in \mathfrak{A}_\infty(T)$, the spectrum of $f(T)$ is exactly the set of values assumed by $f(\lambda)$ as λ varies over the set $\sigma_e(T)$. In symbols, $\sigma[f(T)] = f[\sigma_e(T)]$.*

PROOF. To show that $f[\sigma_e(T)] \subset \sigma[f(T)]$, suppose $\mu \in \sigma(T)$ and define $g(\lambda) = (\lambda - \mu)^{-1}(f(\lambda) - f(\mu))$ if $\lambda \neq \mu$, while $g(\mu) = f'(\mu)$. Then $(\mu - \lambda)g(\lambda) = f(\mu) - f(\lambda)$, and so $(\mu - T)g(T) = f(\mu) - f(T)$, by Lemma 5.6–D. We can permute $\mu - T$ and $g(T)$ on $\mathscr{D}(T)$. It now follows either that the range of $f(\mu) - f(T)$ is not all of X, or that this operator has no inverse; hence $f(\mu) \in \sigma[f(T)]$. It remains to prove that $f(\infty) \in \sigma[f(T)]$ if T is not in $[X]$, or, what is the same thing, if $\mathscr{D}(T) \neq X$. We can assume $f(\infty) - f(\lambda) \neq 0$ if $\lambda \in \sigma(T)$, for otherwise the situation is covered by what has been proved. We can now apply either Theorem 5.6–H or Theorem 5.7–C to the function $f(\infty) - f(\lambda)$ and conclude either that the range of $f(\infty) - f(T)$ is not all of X or that this operator has no inverse. In either event $f(\infty) \in \sigma[f(T)]$.

We now have to prove that $\sigma[f(T)] \subset f[\sigma_e(T)]$. Consider first a point $\beta \in \sigma[f(T)]$ such that $\beta \neq f(\infty)$ and suppose, contrary to what is to be proved, that $f(\lambda) - \beta$ has no zeros on $\sigma(T)$. Define $h(\lambda) = [\beta - f(\lambda)]^{-1}$ on a suitable neighborhood of $\sigma(T)$. Then $h \in \mathfrak{A}_\infty(T)$, and $h(T)$ and $\beta - f(T)$ are inverse to each other, so that $\beta \in \rho[f(T)]$, which is a contradiction. If $\beta = f(\infty)$ and T is not in $[X]$, we know that $\infty \in \sigma_e(T)$, and hence $\beta \in f[\sigma_e(T)]$. It remains only to consider the case $\beta = f(\infty) \in \sigma[f(T)]$, $T \in [X]$. In this case we can replace f by another function g, agreeing with f on a neighborhood of $\sigma(T)$, but such that $g(\infty) \neq \beta$. Then $g(T) = f(T)$ and $g[\sigma_e(T)] = f[\sigma_e(T)]$, and the earlier argument shows that $\beta \in g[\sigma_e(T)]$. This completes the proof.

We also need the following theorem, for the general case of a closed T with nonempty resolvent set. For $T \in [X]$ this is the same as Theorem 5.2–D.

Theorem 5.71–B. *If $P(\lambda)$ is a polynomial, $P[\sigma(T)] = \sigma[P(T)]$.*

The proof can be patterned after that of Theorem 5.71–A. At one stage in the argument the result of § 5.6, problem 7, is needed.

The next theorem increases the scope of the operational calculus by enabling us to deal with composition of functions.

Theorem 5.71–C. *Suppose $f \in \mathfrak{A}_\infty(T)$, $S = f(T)$, and $g \in \mathfrak{A}_\infty(S)$. Suppose also that $f(\infty) \in \Delta(g)$ (if $T \in [X]$ we may always suppose f modified near $\lambda = \infty$, if necessary, to make $f(\infty) \in \Delta(g)$). Define F by $F(\lambda) = g[f(\lambda)]$ if $f(\lambda) \in \Delta(g)$. Then $F \in \mathfrak{A}_\infty(T)$ and $F(T) = g(S)$.*

PROOF. Using Theorem 5.71–A, we see that $\sigma(T) \subset \Delta(F)$. Choose a bounded Cauchy domain D such that $\sigma(S) \subset D$, $\bar{D} \subset \Delta(g)$, and $f(\infty) \in D$. Choose an unbounded Cauchy domain D_1 such that $\sigma(T) \subset D_1$, $\bar{D}_1 \subset \Delta(f)$, and $f(\bar{D}_1) \subset D$. Then

$$F(T) = F(\infty)I + \frac{1}{2\pi i}\int_{+B(D_1)} g[f(\xi)]R_\xi\, d\xi$$

and, if $\xi \in B(D_1)$,

$$g[f(\xi)] = \frac{1}{2\pi i}\int_{+B(D)} \frac{g(\lambda)}{\lambda - f(\xi)}\, d\lambda.$$

If $h(\xi) = [\lambda - f(\xi)]^{-1}$ [λ fixed on $B(D)$], we see that $h(T) = (\lambda - S)^{-1}$. Since

$$F(\infty) = \frac{1}{2\pi i}\int_{+B(D)} \frac{g(\lambda)}{\lambda - f(\infty)}\, d\lambda$$

and

$$g(S) = \frac{1}{2\pi i}\int_{+B(D)} g(\lambda)(\lambda - S)^{-1}\, d\lambda,$$

the conclusion of the theorem follows by easy calculations, which we leave to the reader.

One important application of the preceding theorem is to prove the next theorem, which is used in proving Theorem 5.8–F.

Theorem 5.71–D. *Suppose* $f \in \mathfrak{A}_\infty(T)$, $S = f(T)$, *and let* τ *be a spectral set of* S. *Let* $\sigma = \sigma_e(T) \cap f^{-1}(\tau)$. *Then* σ *is a spectral set of* T, *and the projection* E_σ *associated with* σ *and* T *is the same as the projection* F_τ *associated with* τ *and* S.

PROOF. Let us write $\sigma' = \sigma_e(T) - \sigma$, $\tau' = \sigma(S) - \tau$. From Theorem 5.71–A we see that $f(\sigma \cup \sigma') = \tau \cup \tau'$. Hence $\sigma' = \sigma_e(T) \cap f^{-1}(\tau')$, and it follows that σ and σ' are complementary spectral sets of T. Since $E_\sigma + E_{\sigma'} = I$ and $F_\tau + F_{\tau'} = I$, the relation $E_\sigma = F_\tau$ is implied by $E_{\sigma'} = F_{\tau'}$. Hence, in proving the theorem, it is allowable to assume that σ is a bounded set in the ordinary plane and that $f(\infty)$ is not in τ, for, if this is not true, we can deal instead with σ' and τ'. Now let U_1, U_2, U_3, U_4 be open sets with the following properties: $\bar{U}_1$ and $\bar{U}_2$ are disjoint, and so are $\bar{U}_3$, $\bar{U}_4$; U_1 and U_3 are bounded neighborhoods of τ and σ respectively; U_2 contains τ', $f(\infty)$, and a neighborhood of ∞; U_4 contains σ' and a neighborhood of ∞; $\bar{U}_3$ and $\bar{U}_4$ are in $\Delta(f)$ and $f(\bar{U}_3) \subset U_1$, $f(\bar{U}_4) \subset U_2$. Such sets do exist. Define $f_\tau = 1$ on U_1, $f_\tau = 0$ on U_2, $f_\sigma = 1$ on U_3, $f_\sigma = 0$ on U_4. Then $f_\tau[f(\lambda)] = f_\sigma(\lambda)$ on $U_3 \cup U_4$. But $f_\tau(S) = F_\tau$, $f_\sigma(T) = E_\sigma$, and so $F_\tau = E_\sigma$, by Theorem 5.71–C.

PROBLEMS

1. Suppose $\alpha \in \rho(T)$ and let $A = R_\alpha$. Use Theorem 5.71–A to show that if $\mu(\alpha - \lambda) = 1$, $\mu \in \sigma(A)$ if and only if $\lambda \in \sigma(T)$. If $\mu \in \rho(A)$ and $\mu(\alpha - \beta) = 1$, use Theorem 5.71–C to show that $(\mu - A)^{-1} = \mu^{-1} + \mu^{-2}R_\beta$. Take $f(\lambda) = (\alpha - \lambda)^{-1}$, $g(\lambda) = (\mu - \lambda)^{-1}$. Show also that $R_\beta = \mu A(\mu - A)^{-1}$.

2. Suppose $\alpha \in \rho(T)$ and $A = R_\alpha$. Make the transformation $\mu = (\alpha - \lambda)^{-1}$ from the λ-plane to the μ-plane. If $f \in \mathfrak{A}_\infty(T)$ and g is defined by $g(\mu) = f(\lambda)$, $g(0) = f(\infty)$, show that $g(A) = f(T)$.

3. Suppose X a complex Banach space, A and B in $[X]$, $A \neq B$ and $AB = BA$. Then, if $\lambda_0 \in \sigma(A)$, there exists $\lambda_1 \in \sigma(B)$ with $|\lambda_1 - \lambda_0| \leqslant \|A - B\|$. Suggestion: Suppose the proposition false, and let $C = I - (\lambda_0 - A)(\lambda_0 - B)^{-1} = (A - B)(\lambda_0 - B)^{-1}$. Let $r = \|A - B\|$. Show that $r_\sigma[(\lambda_0 - B)^{-1}] < 1/r$ and hence (see problem 7, § 5.2) that $r_\sigma(C) < 1$, whence $1 \in \rho(C)$. From this follows readily that $\lambda_0 \in \rho(A)$, a contradiction.

4. Consider the operator A of Example 2, § 5.3, with X taken as any fixed l^p. Suppose $\sum_0^\infty |\alpha_n| < \infty$, and let $f(\lambda) = \sum_0^\infty \alpha_n \lambda^n$, $|\lambda| \leqslant 1$. Also let $P_n(\lambda) = \sum_0^n \alpha_k \lambda^k$. Define $S = \sum_0^\infty \alpha_n A^n$ (series convergent in $[X]$). Show that $\sigma(S) = f[\sigma(A)]$. This is not a direct application of Theorem 5.71–A, in general. Why not? For $f[\sigma(A)] \subset \sigma(S)$ use Theorem 4.1–D, and for $\sigma(S) \subset f[\sigma(A)]$ use problem 3 (above) on S and $P_n(A)$.

5.8 Isolated Points of the Spectrum

As in several preceding sections, we assume here that X is a complex Banach space and that T is closed. Then R_λ is analytic as a function of λ on $\rho(T)$, and an isolated point λ_0 of $\sigma(T)$ is an isolated singular point of R_λ. Hence there is a Laurent expansion of R_λ in powers of $\lambda - \lambda_0$. We write this in the form

$$R_\lambda = \sum_0^\infty (\lambda - \lambda_0)^n A_n + \sum_1^\infty (\lambda - \lambda_0)^{-n} B_n. \tag{5.8–1}$$

The coefficients A_n and B_n are members of $[X]$, and this series representation of R_λ is valid when $0 < |\lambda - \lambda_0| < \delta$ for any δ such that *all* of $\sigma(T)$ except λ_0 lies on or outside the circle $|\lambda - \lambda_0| = \delta$. These coefficient operators are given by the usual standard formulas:

$$\begin{aligned} A_n &= \frac{1}{2\pi i}\int_C (\lambda - \lambda_0)^{-n-1} R_\lambda \, d\lambda, \\ B_n &= \frac{1}{2\pi i}\int_C (\lambda - \lambda_0)^{n-1} R_\lambda \, d\lambda, \end{aligned} \tag{5.8–2}$$

where C is any counterclockwise circle $|\lambda - \lambda_0| = h$ with $0 < h < \delta$.

It turns out that there are several important relationships among these coefficient operators. The demonstration of these relationships can be made conveniently by using the operational calculus. Choose $r > 0$ so that $2r < \delta$. Define functions f_n as follows:

$$\begin{aligned} n \geqslant 0: f_n(\lambda) &= \begin{cases} 0 & \text{if } |\lambda - \lambda_0| < r \\ (\lambda - \lambda_0)^{-n-1} & \text{if } |\lambda - \lambda_0| > 2r \end{cases} \\ n < 0: f_n(\lambda) &= \begin{cases} (\lambda - \lambda_0)^{-n-1} & \text{if } |\lambda - \lambda_0| < r \\ 0 & \text{if } |\lambda - \lambda_0| > 2r. \end{cases} \end{aligned} \tag{5.8–3}$$

These functions all belong to $\mathfrak{A}_\infty(T)$, and $f_n(\infty) = 0$. If we compare the definition of $f_n(T)$ with formulas (5.8–2) we see that

$$\begin{aligned} A_n &= -f_n(T), \qquad n \geqslant 0, \\ B_n &= f_{-n}(T), \qquad n \geqslant 1. \end{aligned} \tag{5.8–4}$$

We note in particular that B_1 *is the projection* E_σ *for the case in which* σ *is the spectral set consisting of the single point* λ_0. Since σ is not empty, we know that $B_1 \neq 0$. The only case in which $B_1 = I$ is that in which $\mathscr{D}(T) = X$ and $\sigma(T)$ consists of the single point λ_0. (See the remarks preceding Theorem 5.7–C.)

Now let us observe that

$$\begin{aligned} (\lambda - \lambda_0)f_{n+1}(\lambda) &= f_n(\lambda), && n \geqslant 0, \\ (\lambda - \lambda_0)f_{-n}(\lambda) &= f_{-(n+1)}(\lambda), && n \geqslant 1, \\ (\lambda - \lambda_0)f_0(\lambda) + f_{-1}(\lambda) &= 1. \end{aligned}$$

From these relations and Lemma 5.6–D we obtain the formulas

$$(T - \lambda_0)A_{n+1} = A_n, \tag{5.8–5}$$

$$(T - \lambda_0)B_n = B_{n+1} = (T - \lambda_0)^n B_1, \tag{5.8–6}$$

$$(T - \lambda_0)A_0 = B_1 - I. \tag{5.8–7}$$

The second part of (5.8–6) is obtained from the first part by induction. Some further facts are indicated in problem 1.

As in the classical theory of functions, we shall say that λ_0 is a pole of R_λ of order m if and only if $B_m \neq 0$ and $B_n = 0$ when $n > m$. From (5.8–6) we see that $B_{n+1} = 0$ if $B_n = 0$. Hence λ_0 is a pole of order m if and only if $B_m \neq 0$ and $B_{m+1} = 0$. In that case $B_1, \cdots, B_m$ are all $\neq 0$. If λ_0 is an isolated point of $\sigma(T)$ but not a pole of R_λ, we call it an isolated essential singularity of R_λ.

The case of a pole is of particular interest.

Theorem 5.8–A. *If* λ_0 *is a pole of* R_λ *of order* m, *then* λ_0 *is an eigenvalue of* T. *The ascent and descent of* $\lambda_0 - T$ *are both equal to* m. *The range of the projection* B_1 *is the null space of* $(\lambda_0 - T)^m$, *and the range of* $I - B_1$ *is the range of* $(\lambda_0 - T)^m$, *so that*

$$X = \mathscr{R}(T_0{}^m) \oplus \mathscr{N}(T_0{}^m) \qquad (T_0 = \lambda_0 - T). \tag{5.8–8}$$

The operator T *is completely reduced by the two manifolds occurring in this direct sum.*

PROOF. Let $X_1 = \mathscr{N}(B_1) = \mathscr{R}(I - B_1)$, $X_2 = \mathscr{R}(B_1)$. Then $X = X_1 \oplus X_2$; both subspaces are closed, since B_1 is continuous. We know from the early part of § 5.7 that T is completely reduced by (X_1, X_2), for $I - B_1$ and B_1 are projections associated with complementary spectral sets of T. For convenience we denote the null space and range of $(\lambda_0 - T)^k$ by $\mathscr{N}_k$ and $\mathscr{R}_k$ respectively. If $x \in \mathscr{N}_n$, where $n \geqslant 1$, we see by (5.8–5), induction, and (5.8–7) that $0 = A_{n-1}(T - \lambda_0)^n x =$

$(T - \lambda_0)^n A_{n-1}x = (T - \lambda_0)A_0 x = B_1 x - x$, so that $x = B_1 x \in X_2$. Thus $\mathscr{N}_n \subset X_2$ if $n \geqslant 1$. On the other hand, it follows from (5.8–6) that, if $x \in X_2$, we have $x = B_1 x$ and $(T - \lambda_0)^n x = B_{n+1}x$. Since $B_{n+1} = 0$ if $n \geqslant m$, it follows that $X_2 \subset \mathscr{N}_n$, and hence $\mathscr{N}_n = X_2$, if $n \geqslant m$. But $\mathscr{N}_{m-1}$ is a proper subset of $\mathscr{N}_m$, for $B_m \neq 0$, and $\mathscr{N}_{m-1} = \mathscr{N}_m = X_2$ would imply $B_m = 0$, because of the relation $B_m = (T - \lambda_0)^{m-1}B_1$. We have now proved that the ascent of $\lambda_0 - T$ is m and that $\mathscr{N}_m = X_2$. In particular, since $m > 0$, λ_0 is an eigenvalue of T.

Now let T_1 and T_2 be the restrictions of T to X_1 and X_2 respectively. We know by Theorem 5.7–B*a* that λ_0 is not in $\sigma(T_1)$, for it is in $\sigma(T_2)$. Hence the descent of $\lambda_0 - T_1$ is 0, and the range of $(\lambda_0 - T_1)^n$ is X_1 when $n \geqslant 1$. Hence certainly $X_1 \subset \mathscr{R}_n$. Now, if $n \geqslant m$, the only point common to $\mathscr{R}_n$ and $\mathscr{N}_n$ is 0. For, if $x \in \mathscr{R}_n \cap \mathscr{N}_n$, we can write $x = (\lambda_0 - T)^n y$ and $(\lambda_0 - T)^n x = 0$, whence $y \in \mathscr{N}_{2n} = \mathscr{N}_n$, and therefore $x = 0$. Now suppose $n \geqslant m$ and $x \in \mathscr{R}_n$. Write $x = x_1 + x_2$, where $x_i \in X_i$. Then $x_2 = x - x_1 \in \mathscr{R}_n$, because $X_1 \subset \mathscr{R}_n$. But $x_2 \in X_2 = \mathscr{N}_n$, and so $x_2 = 0$, whence $x = x_1 \in X_1$. Thus $\mathscr{R}_n \subset X_1$ if $n \geqslant m$. We now know that $\mathscr{R}_n = X_1$ if $n \geqslant m$ and, therefore, that the descent of $\lambda_0 - T$ is less than or equal to m. Theorem 5.41–D then shows that the descent is exactly m, which we know to be the ascent. This completes the proof of the theorem.

The next theorem is concerned with conditions on f such that $f(T) = 0$, where $f \in \mathfrak{A}_\infty(T)$. When this theorem is applied to the special case in which X is finite dimensional and $T \in [X]$, it yields the conditions which determine the minimal polynomial associated with T, i.e., the polynomial $P(\lambda)$ of lowest degree such that $P(T) = 0$.

Theorem 5.8–B. *Suppose* $f \in \mathfrak{A}_\infty(T)$. *In order that* $f(T) = 0$ *it is necessary and sufficient that the following conditions be satisfied:*

a. If λ_0 *is a pole of* R_λ *of order* m, *either* $f(\lambda) \equiv 0$ *in some neighborhood of* λ_0 *or* f *has a zero of order at least* m *at* λ_0.

b. $f(\lambda) \equiv 0$ *in some neighborhood of each isolated essential singularity of* R_λ.

c. $f(\lambda) \equiv 0$ *in each component of* $\Delta(f)$ *which contains infinitely many points of* $\sigma(T)$.

d. $f(\lambda) \equiv 0$ *in some neighborhood of* ∞ *if* $\mathscr{D}(T) \neq X$.

PROOF. As a preliminary to the proof let us suppose that λ_0 is an isolated point of $\sigma(T)$, that $f \in \mathfrak{A}_\infty(T)$, and that there is no neighborhood of λ_0 in which $f(\lambda) \equiv 0$. There will then be some smallest integer $k \geqslant 0$ such that $f^{(k)}(\lambda_0) \neq 0$. Now choose r as in (5.8–3), and define g_k so that it is analytic at λ_0 and $g_k(\lambda) = (\lambda - \lambda_0)^{-k} f(\lambda)$ if $0 < |\lambda - \lambda_0| < r$,

$g_k(\lambda) = (\lambda - \lambda_0)^{-k-1}$ if $|\lambda - \lambda_0| > 2r$. We may also restrict r so that f is analytic and $f(\lambda) \neq 0$ when $0 < |\lambda - \lambda_0| < r$. Then $g_k \in \mathfrak{A}_\infty(T)$ and $[g_k(T)]^{-1}$ exists, by Theorem 5.6–H. Referring to (5.8–3), we see that $(\lambda - \lambda_0)^k g_k(\lambda) f_{-1}(\lambda) = f(\lambda) f_{-1}(\lambda)$. Hence $(T - \lambda_0)^k g_k(T) B_1 = f(T) B_1$. We can write this in the forms

$$g_k(T)(T - \lambda_0)^k B_1 = f(T) B_1, \tag{5.8–9}$$

$$B_{k+1} = [g_k(T)]^{-1} f(T) B_1; \tag{5.8–10}$$

to get the second form we make use of (5.8–6).

With these preliminaries established, let us now assume that $f(T) = 0$. If λ_0 is a pole of order m of R_λ and if there is no neighborhood of λ_0 in which $f(\lambda_0) \equiv 0$, we see from (5.8–10) that $B_{k+1} = 0$, where k has the significance explained in the preceding paragraph. Hence, since $B_m \neq 0$, we conclude that $m \leqslant k$. This shows that condition a in the theorem is satisfied. We also see that b is satisfied. For, if λ_0 is an isolated essential singularity of R_λ, then $B_n \neq 0$ for every n, and this is incompatible with (5.8–10) if $f(T) = 0$. Next we observe that $\sigma[f(T)]$ consists of the single point 0 when $f(T) = 0$. Since $\sigma[f(T)] = f[\sigma_e(T)]$, we conclude that $f(\lambda) = 0$ at all points of $\sigma_e(T)$. That condition c must be satisfied now follows from the fact that f is locally analytic on $\Delta(f)$ and at ∞, for $\sigma_e(T)$ is closed and the extended plane is compact. Finally, suppose $\mathscr{D}(T) \neq X$. Then $f(\infty) = 0$, since $\infty \in \sigma_e(T)$. If ∞ is an accumulation point of $\sigma(T)$, it follows that $f(\lambda) \equiv 0$ in the component of $\Delta(f)$ which contains ∞. If ∞ is isolated in $\sigma_e(T)$ and $f(\lambda)$ is not identically zero in any neighborhood of ∞, this means that ∞ is a zero of f of some finite order k. Let $g(\lambda) = \lambda^k f(\lambda)$. Then $g \in \mathfrak{A}_\infty(T)$ and $g(\infty) \neq 0$. But $g(T) = T^k f(T) = 0$, and therefore $g(\infty) = 0$, for $\sigma[g(T)] = (0) = g[\sigma_e(T)]$. Thus ($d$) must be satisfied if $f(T) = 0$.

Now, conversely, suppose that f satisfies conditions a–d in the theorem; we shall show that $f(T) = 0$. We may suppose that there is at least one component of $\Delta(f)$ in which $f(\lambda) \not\equiv 0$, since otherwise the conclusion $f(T) = 0$ is a direct consequence of the defining formula (5.6–10). Because of the conditions on f, the fact that $\sigma(T)$ is closed and $\sigma(T) \subset \Delta(f)$, it follows that there are at most a finite number of points of $\sigma(T)$, say $\lambda_1, \cdots, \lambda_p$, all of them poles of R_λ, in the union of those components of $\Delta(f)$ where $f(\lambda) \not\equiv 0$. Now, if $\mathscr{D}(T) \neq X$, condition d shows that the definition of $f(T)$ can be put in the form

$$f(T) = \sum_{j=1}^{p} \frac{1}{2\pi i} \int_{C_j} f(\lambda) R_\lambda \, d\lambda, \tag{5.8–11}$$

where $C_1, \cdots, C_p$ are small nonoverlapping counterclockwise circles with centers at $\lambda_1, \cdots, \lambda_p$. If no such points as $\lambda_1, \cdots, \lambda_p$ exist, then $f(T) = 0$ by Cauchy's theorem. If $\mathscr{D}(T) = X$, formula (5.8–11) is still valid, for the behavior of f near ∞ is irrelevant in this case. Let E_j be the projection associated with the spectral set consisting of the single point λ_j. It is then clear by the definition of E_j and the homomorphism rules of the operational calculus that (5.8–11) becomes

$$f(T) = \sum_{j=1}^{p} f(T)E_j.$$

It therefore remains only to prove that $f(T)E_j = 0$ for each j. Let λ_j be a pole of order m_j of R_λ, and let its order as a zero of f be k_j, so that $m_j \leqslant k_j$, by condition *a*. If we apply (5.8–6) with λ_j in place of λ_0 and E_j in place of B_1, the fact that λ_j is a pole of order m_j implies $(T - \lambda_j)^{k_j}E_j = 0$. Formula (5.8–9), applied with this change in notation, shows that $f(T)E_j = 0$. This finishes the proof.

Our next principal result is Theorem 5.8–D, which is a sort of converse of Theorem 5.8–A. An important tool in the proof of this theorem is furnished by the following lemma.

Lemma 5.8–C. *Let λ_0 be an isolated point of $\sigma(T)$, and let B_1 be the projection associated with λ_0 as a spectral set. Then $x \in \mathscr{R}(B_1)$ if and only if $\lim_{n\to\infty} (T - \lambda_0)^n\epsilon^{-n}x = 0$ for every $\epsilon > 0$ (or, equivalently, if and only if $\|(T - \lambda_0)^n x\|^{1/n} \to 0$). In order that this be so it is sufficient to have it true for a single ϵ such that $\lambda \in \rho(T)$ if $0 < |\lambda - \lambda_0| \leqslant \epsilon$.*

PROOF. If $x \in \mathscr{R}(B_1)$ we have $x = B_1x$. It follows from (5.8–6) that $B_{n+1}x = (T - \lambda_0)^n x$. Then, by (5.8–2),

$$(T - \lambda_0)^n\epsilon^{-n}x = \frac{1}{2\pi i}\int_C (\lambda - \lambda_0)^n\epsilon^{-n}R_\lambda x \, d\lambda,$$

where the radius of C may be taken less than ϵ. Consequently, $(T - \lambda_0)^n\epsilon^{-n}x \to 0$. For the converse the motivation is that of finding an operator which is nearly the same as B_1 but has an inverse. We define

$$h_n(\lambda) = \{1 - (\lambda - \lambda_0)^n\epsilon^{-n}\}^{-1},$$

where ϵ is prescribed as in the last sentence of the lemma. Then Theorem 5.6–H applies to h_n; $h_n(T)$ has an inverse, which is evidently $I - (T - \lambda_0)^n\epsilon^{-n}$. To calculate $h_n(T)$ as an integral, let C_1 be the circle $|\lambda - \lambda_0| = r_1$, C_2 the circle $|\lambda - \lambda_0| = r_2$, where $0 < r_1 < \epsilon < r_2$ and

$\lambda \in \rho(T)$ if $0 < |\lambda - \lambda_0| \leqslant r_2$. Then, with proper orientation of the circles,

$$h_n(T) = \frac{1}{2\pi i} \int_{C_1} h_n(\lambda) R_\lambda \, d\lambda - \frac{1}{2\pi i} \int_{C_2} h_n(\lambda) R_\lambda \, d\lambda.$$

We can also write

$$B_1 = \frac{1}{2\pi i} \int_{C_1} R_\lambda \, d\lambda.$$

From these formulas it is easy to see that $\|h_n(T) - B_1\| \to 0$. If now $(T - \lambda_0)^n \epsilon^{-n} x \to 0$, we can write $x = h_n(T)[I - (T - \lambda_0)^n \epsilon^{-n}]x$, and so

$$x - B_1 x = [h_n(T) - B_1]x - h_n(T)(T - \lambda_0)^n \epsilon^{-n} x \to 0.$$

Thus $x = B_1 x \in \mathscr{R}(B_1)$, as was to be proved.

In the next theorem we consider a fixed λ_0, and we denote the range and null space of $(\lambda_0 - T)^n$ by $\mathscr{R}_n$ and $\mathscr{N}_n$ respectively.

Theorem 5.8–D. *Suppose there is an integer n such that* $\mathscr{R}_n \cap \mathscr{N}_n = (0)$, *that* $\mathscr{R}_n$ *is closed and* $X = \mathscr{R}_n \oplus \mathscr{N}_n$, *and that the smallest integer for which all these things are true is* m, *with* $m \geqslant 1$. *Then* λ_0 *is a pole of* R_λ *of order* m.

PROOF. We observe at the outset that $\mathscr{N}_m$ contains nonzero elements. For, if $\mathscr{N}_m = (0)$, then $\mathscr{N}_{m-1} = (0)$ also. Then $X = \mathscr{R}_m \oplus \mathscr{N}_m = \mathscr{R}_m \subset \mathscr{R}_{m-1}$, whence $X = \mathscr{R}_{m-1}$. Then $\mathscr{R}_{m-1} \cap \mathscr{N}_{m-1} = (0)$, $\mathscr{R}_{m-1}$ is closed, and $X = \mathscr{R}_{m-1} \oplus \mathscr{N}_{m-1}$, contrary to the definition of m.

First we shall show that T is completely reduced by $(\mathscr{R}_m, \mathscr{N}_m)$. We leave to the reader the easy verification that $T\mathscr{N}_m \subset \mathscr{N}_m$ and $T[\mathscr{R}_m \cap \mathscr{D}(T)] \subset \mathscr{R}_m$. If we write $x = x_1 + x_2$, where $x_1 \in \mathscr{R}_m$, $x_2 \in \mathscr{N}_m$, let $E_1 x = x_1$, $E_2 x = x_2$. The projections E_1, E_2 are continuous, by Theorem 4.8–D. Clearly $\mathscr{D}(T)$ is invariant under E_2, since $\mathscr{N}_m \subset \mathscr{D}(T)$. Also, $\mathscr{D}(T)$ is invariant under E_1, because $x \in \mathscr{D}(T)$ implies $E_1 x = x - E_2 x \in \mathscr{D}(T)$. This finishes the demonstration that T is completely reduced by $(\mathscr{R}_m, \mathscr{N}_m)$. Let T_1 and T_2 denote the restrictions of T to $\mathscr{R}_m$ and $\mathscr{N}_m$ respectively. Then T_2 is defined on all of $\mathscr{N}_m$ and is therefore continuous, for T_2 is closed. Also, Theorem 5.4–C is applicable here, and in particular, $\sigma(T) = \sigma(T_1) \cup \sigma(T_2)$.

The next step is to show that λ_0 is a pole of the resolvent of T_2. Choose a fixed point $\alpha \in \rho(T)$ [then $\alpha \in \rho(T_2)$ also], and let $f(\lambda) = (\lambda_0 - \lambda)^m (\alpha - \lambda)^{-m-1}$. Then $f(T) = (\lambda_0 - T)^m R_\alpha^{m+1}$. Now f is in $\mathfrak{A}_\infty(T_2)$ as well as in $\mathfrak{A}_\infty(T)$, and it is easily verified that $f(T_2)$ is the restriction of $f(T)$ to $\mathscr{N}_m$. But, if $x \in \mathscr{N}_m$, $f(T)x = R_\alpha^{m+1}(\lambda_0 - T)^m x = 0$, and so $f(T_2) = 0$. We now apply Theorem 5.8–B. The conclusion is that $\sigma(T_2)$ consists of the single point λ_0, which is a pole of some order m_1 of $(\lambda - T_2)^{-1}$, with $1 \leqslant m_1 \leqslant m$.

Next we prove that $\lambda_0 \in \rho(T_1)$. From this it will follow that λ_0 is an isolated point of $\sigma(T)$. Let $P(\lambda) = (\lambda_0 - \lambda)^m$. We shall show that $P(T_1) = (\lambda_0 - T_1)^m$ has an inverse with domain all of $\mathscr{R}_m$. If $x \in \mathscr{R}_m \cap \mathscr{D}_m(T)$ and $P(T_1)x = 0$, this means that $x \in \mathscr{R}_m \cap \mathscr{N}_m$, and hence that $x = 0$. Therefore $P(T_1)$ has an inverse. If $y \in \mathscr{R}_m$, we can write $y = (\lambda_0 - T)^m x$ for some $x \in \mathscr{D}_m(T)$, by definition. If we write $x = E_1 x + E_2 x$, it follows that $y = (\lambda_0 - T)^m E_1 x = P(T_1)E_1 x$. Thus the range of $P(T_1)$ is all of $\mathscr{R}_m$. Now $P(T_1)$ is closed (see problem 4, § 5.6), and so the inverse of $P(T_1)$ is a closed operator defined on all of $\mathscr{R}_m$; it is therefore continuous, by the closed-graph theorem. This means that $0 \in \rho[P(T_1)]$. But $\sigma[P(T_1)] = P[\sigma(T_1)]$ (Theorem 5.71–B). Consequently $P(\lambda)$ has no zeros on $\sigma(T_1)$; i.e., λ_0 is not in $\sigma(T_1)$.

Now we can regard the single point λ_0 as a spectral set of T and obtain the corresponding projection B_1. We shall prove that $\mathscr{R}(B_1) = \mathscr{N}_m$. For this we use Lemma 5.8–C. It is clear that $\mathscr{N}_m \subset \mathscr{R}(B_1)$, for $(\lambda_0 - T)^n \epsilon^{-n} x = 0$ if $\epsilon > 0$, $n \geqslant m$, and $x \in \mathscr{N}_m$. Let $H = (\lambda_0 - T_1)^{-1}$. *We have just seen that* $\lambda_0 \in \rho(T_1)$. *Choose* $\epsilon > 0$ *so that* $\epsilon \|H\| < 1$, and suppose $x \in \mathscr{R}(B_1)$. Then $(\lambda_0 - T)^n \epsilon^{-n} x \to 0$. Now certainly $(\lambda_0 - T)^n E_2 x = 0$ if $n \geqslant m$, and hence $(\lambda_0 - T)^n \epsilon^{-n} E_1 x \to 0$. Let $y_n = (\lambda_0 - T)^n \epsilon^{-n} E_1 x$. Then $y_n \in \mathscr{R}_m$ and $H^n y_n = \epsilon^{-n} E_1 x$. Hence $\|E_1 x\| = \|\epsilon^n H^n y_n\| \leqslant (\epsilon \|H\|)^n \|y_n\|$. We conclude that $E_1 x = 0$, whence $x \in \mathscr{N}_m$. Thus $\mathscr{R}(B_1) = \mathscr{N}_m$.

Now, the fact that λ_0 is a pole of order m_1 of $(\lambda - T_2)^{-1}$, the underlying space being $\mathscr{N}_m$, enables us to conclude that λ_0 is also a pole of order m_1 of R_λ. The details of this inference are indicated in problem 2 at the end of the section. It remains only to show that $m = m_1$. We know that $m_1 \leqslant m$; $m_1 < m$ is impossible, as we see by Theorem 5.8–A and the assumed minimal character of m. Thus, finally, the proof of Theorem 5.8–D is complete.

We now have the following result:

Theorem 5.8–E. *If T belongs to $[X]$ and is compact, each nonzero point of $\sigma(T)$ is a pole of R_λ.*

PROOF. Suppose $\lambda_0 \in \sigma(T)$, $\lambda_0 \neq 0$. Then $\alpha(\lambda_0 - T) = \delta(\lambda_0 - T) = m \geqslant 1$ (Theorem 5.5–E). The fact that λ_0 is a pole of order m of R_λ now follows by Theorems 5.5–D, 5.41–G, 5.8–D, 5.8–A. Detailed verification is left to the reader.

A different proof of this theorem can be had by way of the next theorem.

Since a pole of R_λ is an eigenvalue of T (Theorem 5.8–A), it is useful to have some means of deciding whether a point of $\sigma(T)$ is a pole of R_λ. One important result along such lines is contained in the following theorem, in which T can be any closed operator with nonempty resolvent set.

Theorem 5.8–F. *Suppose* $f \in \mathfrak{A}_\infty(T)$, *and let* $f(T)$ *be compact. Then, if* $\lambda_0 \in \sigma(T)$ *and* $f(\lambda_0) \neq 0$, λ_0 *is a pole of* R_λ. *Moreover, if* B_1 *is the projection associated with the spectral set* (λ_0), *the range of* B_1 *is finite dimensional, and hence the eigenmanifold corresponding to* λ_0 *is finite dimensional.*

PROOF. Let $S = f(T)$, $\mu = f(\lambda_0)$. We know that $\mu \in \sigma(S)$, by Theorem 5.71–A. Let $\sigma = \sigma_e(T) \cap f^{-1}(\mu)$. We know from § 5.5 that μ is isolated in $\sigma(S)$, so that (μ) is a spectral set of S. Then σ is a spectral set of T, and the projection associated with (μ) and S is E_σ (associated with σ and T), by Theorem 5.71–D. Let $X_\sigma = \mathscr{R}(E_\sigma)$, and let S_0 be the restriction of S to X_σ. Then $\sigma(S_0) = (\mu)$ (Theorem 5.7–B), and so $0 \in \rho(S_0)$, whence $x = S_0 S_0^{-1} x$ if $x \in X_\sigma$. Since S_0 is compact, this implies (by means of Theorem 3.12–F) that X_σ is finite dimensional. Now let T_1 be the restriction of T to X_σ. We know from Theorem 5.7–B that $\sigma = \sigma_e(T_1)$. We also know that $\sigma_e(T_1)$ cannot be the entire extended plane, for $\sigma_e(T_1) \subset \sigma_e(T)$ and $\rho(T)$ is not empty, by assumption. It then follows that $X_\sigma \subset \mathscr{D}(T)$, for otherwise we could conclude from the finite dimensionality of X_σ that for no λ can the range of $\lambda - T_1$ be all of X_σ and hence that $\sigma_e(T_1)$ is the entire extended plane. Consequently $T_1 \in [X_\sigma]$ and $\sigma = \sigma(T_1)$ is a finite set. Since $\lambda_0 \in \sigma$, this means that λ_0 is isolated in $\sigma(T)$. It is easily seen that $E_\sigma = B_1 + P$, $PB_1 = 0$, where P is the projection associated with $\sigma - (\lambda_0)$ and B_1 is defined as in the theorem. Hence $\mathscr{R}(B_1) \subset X_\sigma$, and so $\mathscr{R}(B_1)$ is finite dimensional. When T is restricted to $\mathscr{R}(B_1)$, its resolvent must have a pole at λ_0, because of the finite dimensionality. But then λ_0 is a pole of R_λ (see problem 2). Since the eigenmanifold corresponding to λ_0, namely $\mathscr{N}(\lambda_0 - T)$, is contained in $\mathscr{R}(B_1)$, it must be finite dimensional.

The foregoing proof uses certain facts about the spectrum of an operator in a finite-dimensional space. These facts are evident as soon as we introduce a basis and represent the operator as a matrix, using determinants to find the spectrum and the resolvent. A discussion which covers these matters without appeal to determinants appears in § 5.9.

One important application of Theorem 5.8–F is to the case in which $T \in [X]$ and T^n is compact for some positive integer n. Clearly we can choose $f \in \mathfrak{A}_\infty(T)$ so that $f(\lambda) = \lambda^n$ on a neighborhood of $\sigma(T)$, and then $f(T) = T^n$. *Hence, if* T^n *is compact for some* n, *every nonzero point in* $\sigma(T)$ *is a pole of* R_λ, *and the corresponding eigenmanifold is finite dimensional.* It can happen, for example, that T^2 is compact though T is not.

Example. Take $X = L(a, b)$, where (a, b) is a finite or infinite interval. Let $k(s, t)$ be a measurable function on $(a, b) \times (a, b)$ such that $|k(s, t)| \leqslant h(s)$ almost everywhere, where $h \in X$. Define $Kx = y$ to mean $y(s) = \int_a^b k(s, t)x(t)\,dt$. Then $K \in [X]$; K need not be compact, but K^2 is compact. See Zaanen, **1**, pages 322–323. Also Dunford and Pettis, **1**, page 370 and Phillips, **1**, page 536.

Another important application is to the case of an operator (not necessarily a member of $[X]$) whose resolvent is compact. We note that as a result of (5.1–2), if R_λ is compact for one λ in $\rho(T)$, it is compact for every such λ (see problems 3 and 4 in § 5.5). Suppose $\alpha \in \rho(T)$ and that R_α is compact. Now $R_\alpha = f(T)$, where $f(\lambda) = (\alpha - \lambda)^{-1}$. *Theorem* 5.8–F *then shows that in this case every point of* $\sigma(T)$ *is a pole of* R_λ. This is the situation which prevails in the case of certain differential operators, that of Example 4, § 5.6, for instance.

PROBLEMS

1. Show that the coefficients in (5.8–1) satisfy the relations (*a*) $A_m A_n = -A_{m+n+1}$, (*b*) $B_m B_n = B_{m+n-1}$, (*c*) $A_m B_n = 0$, (*d*) $A_n = (-1)^n A_0^{n+1}$. Show also that $\mathscr{R}(A_n) \subset \mathscr{D}_{n+1}(T)$ if $n \geqslant 0$ and $\mathscr{R}(B_n) \subset \mathscr{D}_k(T)$ if $n \geqslant 1$ and $k \geqslant 0$.

2. Let λ_0 be an isolated point of $\sigma(T)$, and consider the Laurent expansion (5.8–1). Show that $\sum_1^\infty (\lambda - \lambda_0)^{-n} B_n$ converges if $\lambda \neq \lambda_0$. Let $X_0 = \mathscr{R}(B_1)$, and let T_0 be the restriction of T to X_0. Then $T_0 \in [X_0]$. Let C_n be the restriction of B_n to X_0, and let $S_\lambda = \sum_1^\infty (\lambda - \lambda_0)^{-n} C_n$. Using (5.8–6), show that $S_\lambda = (\lambda - T_0)^{-1}$ if $\lambda \neq \lambda_0$. Note that $B_n B_1 = B_n$ and so conclude that λ_0 is a pole of order m of R_λ if and only if it is a pole of order m of S_λ.

3. Take $X = l^1$ and define $A \in [X]$ by the matrix (α_{ij}), where $\alpha_{ii} = \beta_i$, $\alpha_{12} = \alpha_{34} = \alpha_{56} = \cdots = 1$, and all other entries in the matrix are 0. Suppose $\beta_i \to 0$. Then A is not compact, but A^2 is compact. Discuss $\sigma(A)$ and the resolvent.

4. If $\mathscr{D}(T) \neq X$ there exists no polynomial $P(\lambda) \neq 0$ such that $P(T)x = 0$ for every $x \in \mathscr{D}_n(T)$ (n the degree of P). If such a polynomial P did exist, choose $\alpha \in \rho(T)$, $f(\lambda) = (\alpha - \lambda)^{-n-1} P(\lambda)$, and show that $f(T) = 0$. Then use Theorem 5.8–B.

5. Suppose that $\sigma(T)$ is bounded and that $\mathscr{D}(T) \neq X$. Then R_λ has an expansion in powers of λ valid when $|\lambda| > r_\sigma(T)$, say $R_\lambda = \sum_1^\infty \lambda^{-n} A_n + \sum_0^\infty \lambda^n B_n$. Show that $TB_0 + I = A_1$, $TB_n = B_{n-1}$, $TA_n = A_{n+1}$. Show that $B_n \neq 0$ for each n and that A_1 is the projection associated with the spectral set $\sigma(T)$.

5.9 Operators with Rational Resolvent

Let X be a complex Banach space. A function f of the complex variable λ, with values in X, is called a rational function if it is expressible in the form $f(\lambda) = p(\lambda)/Q(\lambda)$, where $p(\lambda)$ is a polynomial in λ with coefficients in X and $Q(\lambda)$ is a polynomial with complex coefficients. We assume that $p(\lambda)$ and $Q(\lambda)$ are never zero at the same point. A rational function also has a partial-fractions representation. Just as in classical analysis it may be proved with the aid of Liouville's theorem that, if the function f is holomorphic on the whole plane except for a finite number of poles and if $f(1/\lambda)$ has a pole or a removable singularity at $\lambda = 0$, then f is rational. This is done by expressing $f(\lambda)$ as the sum of a polynomial and the singular parts of its Laurent expansions at the various poles.

In this section we consider operators T for which the resolvent R_λ is a rational function with values in $[X]$. If $T \in [X]$ and $\sigma(T)$ consists of a finite number of points, each a pole of R_λ, then R_λ is rational, for $\|R_\lambda\| \to 0$ as $\lambda \to \infty$. It is impossible to have R_λ rational if T is closed but not in $[X]$ (see problem 5, § 5.8, which indicates that R_λ has an essential singularity at ∞ in this case). Our discussion of operators with rational resolvents includes a determinant-free treatment of the situation when X is finite dimensional.

We begin with consideration of the question as to when $P(T) = 0$ if $T \in [X]$ and $P(\lambda)$ is a scalar polynomial. We could appeal to Theorem 5.8–B, but we prefer a more elementary argument.

Theorem 5–9A. *Suppose $T \in [X]$, and let $P(\lambda)$ be a scalar polynomial of degree $\geqslant 1$. Then $P(T) = 0$ if and only if there exists a polynomial $q(\lambda)$ with coefficients in $[X]$ such that*

$$P(\lambda)R_\lambda = q(\lambda) \tag{5.9–1}$$

when $\lambda \in \rho(T)$.

PROOF. Let $Q(\lambda, \mu)$ be the polynomial in λ and μ such that $P(\lambda) - P(\mu) \equiv (\lambda - \mu)Q(\lambda, \mu)$. With T in place of μ we have $P(\lambda) - P(T) = Q(\lambda, T)(\lambda - T)$. Multiplication by R_λ gives $P(\lambda)R_\lambda - P(T)R_\lambda = Q(\lambda, T)$. Suppose now that $P(T) = 0$. Then we obtain (5.9–1), with $q(\lambda) = Q(\lambda, T)$. On the other hand, if we start from (5.9–1), we see that $P(T)R_\lambda = q(\lambda) - Q(\lambda, T)$, and hence

$$\|q(\lambda) - Q(\lambda, T)\| \leqslant \|P(T)\| \, \|R_\lambda\|.$$

Since $\|R_\lambda\| \to 0$ as $\lambda \to \infty$, we conclude by Liouville's theorem that

$q(\lambda) - Q(\lambda, T)$ is a polynomial in λ which is identically zero. It follows that $P(T)R_\lambda = 0$, and hence $P(T) = P(T)R_\lambda(\lambda - T) = 0$.

Theorem 5.9–B. *Suppose $T \in [X]$. There exists a scalar polynomial $P(\lambda)$ of degree $\geqslant 1$ such that $P(T) = 0$ if and only if R_λ is rational.*

PROOF. If R_λ is a rational function of λ, it is expressible in the form $R_\lambda = q(\lambda)/P(\lambda)$, where q and P are polynomials with no common zeros. We know that $\sigma(T)$ is not empty (Theorem 5.2–B); hence the degree of $P(\lambda)$ is at least 1, for $\sigma(T)$ is precisely the set of singular points of R_λ, i.e., the set of zeros of $P(\lambda)$. It now follows from Theorem 5.9–A that $P(T) = 0$. Suppose, on the other hand, that $P(\lambda)$ is some polynomial of degree $\geqslant 1$ such that $P(T) = 0$. By Theorem 5.9–A we know there exists a polynomial $q(\lambda)$ (with coefficients in $[X]$) such that (5.9–1) holds when $\lambda \in \rho(T)$. It remains to show that there exist polynomials P_1 and q_1 such that $\rho(T) = \{\lambda : P_1(\lambda) \neq 0\}$, P_1 and q_1 have no common zeros, and $R_\lambda = q_1(\lambda)/P_1(\lambda)$. It is not necessarily true that we can take $P_1 = P$ and $q_1 = q$.

If $\lambda_0 \in \rho(T)$ and if λ_0 is a common zero of P and q, its multiplicity m as a zero of P is the same as its multiplicity n as a zero of q. For, if we let $r = \min(m, n)$, divide both sides of (5.9–1) by $(\lambda - \lambda_0)^r$, and let $\lambda \to \lambda_0$, we get a contradiction unless $m = n$ (observe that $R_\lambda \neq 0$ when $\lambda \in \rho(T)$, as a result of our standing assumption that X has at least two elements). Now let P_0 and q_0 be the polynomials obtained from P and q respectively, by removal of the factors arising from common zeros in $\rho(T)$. Then $P_0(\lambda)$ has no zeros in $\rho(T)$ and

$$R_\lambda = \frac{q_0(\lambda)}{P_0(\lambda)} \tag{5.9–2}$$

if $\lambda \in \rho(T)$.

Next we show that $\sigma(T)$ is precisely the set of zeros of $P_0(\lambda)$. If $\pi = \{\lambda : P_0(\lambda) \neq 0\}$, we know that $\rho(T) \subset \pi$. Now π is connected and $\rho(T)$ is open, so that, if $\pi - \rho(T)$ were nonempty, it would have to contain an accumulation point of $\rho(T)$, say α. Let $A = q_0(\alpha)/P_0(\alpha)$. If $\{\lambda(n)\}$ is a sequence in $\rho(T)$ such that $\lambda(n) \to \alpha$, we see from (5.9–2) that $R_{\lambda(n)} \to A$. From this we easily conclude that $A(\alpha - T) = (\alpha - T)A = I$, so that $\alpha \in \rho(T)$ and $A = R_\alpha$. We must therefore have $\pi = \rho(T)$.

Now, finally, there may be common zeros of P_0 and q_0 *not* in $\rho(T)$. However, the multiplicity of such a zero for P_0 must exceed its multiplicity for q_0, for a contrary assumption would lead us to the conclusion that the zero is in $\rho(T)$, by an argument like the foregoing. For each such zero β, of multiplicity n as a zero of q_0, let us divide $P_0(\lambda)$ and $q_0(\lambda)$ by $(\lambda - \beta)^n$. Let $P_1(\lambda)$ and $q_1(\lambda)$ be the polynomials which result from

$P_0(\lambda)$ and $q_0(\lambda)$ after the elimination of all such factors $(\lambda - \beta)^n$. Then P_1 and q_1 have no common zeros, $\sigma(T)$ is precisely the set of zeros of P_1, and $R_\lambda = q_1(\lambda)/P_1(\lambda)$. This shows that R_λ is a rational function, and completes the proof of the theorem.

The foregoing proof shows that, if $P(T) = 0$, then $P(\lambda)$ is divisible by the denominator $P_1(\lambda)$ of the rational function R_λ. Hence, if R_λ is rational, with distinct poles $\lambda_1, \cdots, \lambda_k$ of orders $m_1, \cdots, m_k$, among all polynomials $P(\lambda)$ of degree $\geqslant 1$ such that $P(T) = 0$ there is a unique one of lowest degree with leading coefficient 1, and this unique one is

$$(\lambda - \lambda_1)^{m_1} \cdots (\lambda - \lambda_k)^{m_k}.$$

This polynomial is called the *minimal polynomial* associated with T. We recall from Theorem 5.8–A that m_i is the ascent of $\lambda_i - T$.

For finite-dimensional spaces we have:

Theorem 5.9–C. *Suppose X is of finite dimension $n(n \geqslant 1)$ and that $T \in [X]$. Then R_λ is rational.*

PROOF. We deliberately avoid the use of determinants. Let us write $T(\lambda) = \lambda - T$. The ascent and descent of $T(\lambda)$ must be finite, because of the finite dimensionality. They are equal, by Theorem 5.41–E. If the ascent is 0, then $\lambda \in \rho(T)$; otherwise $\lambda \in P\sigma(T)$. Now, eigenvectors corresponding to distinct eigenvalues are linearly independent (see the first part of the proof of Theorem 5.5–G). Hence, since X is n-dimensional, $\sigma(T)$ cannot contain more than n points. We know that $\sigma(T)$ is not empty. Let the distinct points of $\sigma(T)$ be $\lambda_1, \cdots, \lambda_k$, and let the ascent of $T(\lambda_i)$ be m_i. We define $P(\lambda) = (\lambda - \lambda_1)^{m_1} \cdots, (\lambda - \lambda_k)^{m_k}$. Select any $x \neq 0$. Since X is n-dimensional, there exist scalars $\alpha_0, \alpha_1, \cdots, \alpha_n$ (perhaps depending on x) not all 0, such that $\alpha_0 x + \alpha_1 Tx + \cdots + \alpha_n T^n x = 0$. Let $Q(\lambda) = \alpha_0 + \alpha_1\lambda + \cdots + \alpha_n\lambda^n$, and let the factored form of Q be $\alpha(\lambda - \beta_1)^{\nu_1} \cdots (\lambda - \beta_r)^{\nu_r}$, where $\beta_1, \cdots, \beta_r$ are the distinct zeros, and $\alpha \neq 0$. Then

$$(T - \beta_1)^{\nu_1} \cdots (T - \beta_r)^{\nu_r} x = 0. \tag{5.9–3}$$

From this we propose to conclude that $P(T)x = 0$. For any β_i that is not one of the points $\lambda_1, \cdots, \lambda_k$ we can omit the factor $(T - \beta_i)^{\nu_i}$ from (5.9–3) without impairing the equality, because $(T - \beta_i)^{\nu_i}$ has an inverse in this case. On the other hand, for any λ_j that is not one of the points $\beta_1, \cdots, \beta_r$ we can insert the factor $(T - \lambda_j)^{m_j}$ in (5.9–3) without destroying the equality. It remains to consider the case of a β_i which *is* one of the points $\lambda_1, \cdots, \lambda_k$. For simplicity of notation suppose $\beta_1 = \lambda_1$. If $\nu_1 < m_1$ we can evidently replace ν_1 by m_1 in (5.9–3), and the equation will still hold. The same is true if $m_1 < \nu_1$, the reason being that

$\mathcal{N}[T(\beta_1)^{\nu_1}] = \mathcal{N}[T(\beta_1)^{m_1}]$ in this case, because m_1 is the ascent of $\beta_1 - T$. It is now clear that $P(T)x = 0$. But then $P(T) = 0$, since $P(T)$ is independent of x and x was arbitrary. Hence R_λ is rational, by Theorem 5.9–B.

We now turn to some general considerations which will lead us to a result about reducibility in the case of operators for which R_λ is rational. We shall obtain in a more elementary way a result which is obtainable from Theorems 5.7–A and 5.8–A. In the immediately following theorem no topology is needed.

Theorem 5.9–D. *Let X be any complex linear space, and suppose T is a linear operator on X into X. (a) Let $P_1(\lambda)$ and $P_2(\lambda)$ be scalar polynomials without common zeros. Let $P(\lambda) = P_1(\lambda)P_2(\lambda)$, and let M_1, M_2, M be the null manifolds of $P_1(T)$, $P_2(T)$, $P(T)$ respectively. Then $M_1 \cap M_2 = (0)$ and $M = M_1 \oplus M_2$. (b) Let $\lambda_1, \cdots, \lambda_k$ be distinct complex numbers, and let $m_1, \cdots, m_k$ be positive integers. Let $P(\lambda) = (\lambda - \lambda_1)^{m_1} \cdots (\lambda - \lambda_k)^{m_k}$. Let M_i be the null manifold of $(T - \lambda_i)^{m_i}$, and let M be the null manifold of $P(T)$. Then $M_1, \cdots, M_k$ are linearly independent, and $M = M_1 \oplus \cdots \oplus M_k$.*

PROOF. (*a*) Since P_1 and P_2 are relatively prime, there exist polynomials Q_1 and Q_2 such that

$$Q_1(\lambda)P_1(\lambda) + Q_2(\lambda)P_2(\lambda) \equiv 1$$

and hence

$$Q_1(T)P_1(T) + Q_2(T)P_2(T) = I. \tag{5.9–4}$$

It follows that

$$Q_1(T)[P_1(T)]^2 + Q_2(T)P(T) = P_1(T). \tag{5.9–5}$$

From (5.9–4) we see that $M_1 \cap M_2 = (0)$. Now suppose $x \in M$, and let $y = Q_1(T)P_1(T)x$. Then (5.9–5) shows that $P_1(T)y = P_1(T)x$, or $y - x \in M_1$. Also, $P_2(T)y = Q_1(T)P(T)x = 0$, so that $y \in M_2$. Then $x = (x - y) + y$ shows that $M \subset M_1 \oplus M_2$. It is evident that $M_1 \subset M$ and $M_2 \subset M$, and hence $M = M_1 \oplus M_2$. This finishes (*a*). The truth of (*b*) follows from (*a*) by induction.

We now revert to the case in which X is a complex Banach space and $T \in [X]$.

Theorem 5.9–E. *Suppose R_λ is rational, with $P(\lambda) = (\lambda - \lambda_1)^{m_1} \cdots (\lambda - \lambda_k)^{m_k}$ the minimal polynomial for T. Then, in the notation of* Theorem 5.9–D*b*, *$X = M_1 \oplus \cdots \oplus M_k$, and T is completely reduced by $(M_1, \cdots, M_k)$.*

PROOF. The fact that $T(M_i) \subset M_i$ is evident from $(T - \lambda_i)^{m_i}T = T(T - \lambda_i)^{m_i}$, and $M = X$, because $P(T) = 0$.

The direct-sum decomposition of X described in Theorem 5.9–E determines projections E_i defined by $E_i x = x_i$ if $x = x_1 + \cdots + x_k$, $x_i \in M_i$. The projection E_i may be described as the residue of R_λ at the pole λ_i; it is the projection associated with the spectral set (λ_i). Let

$$N = \sum_{i=1}^{k} (T - \lambda_i)E_i,$$

and let $p = \max(m_1, \cdots, m_k)$. It is easily seen that

$$N^p = \sum_{i=1}^{k} (T - \lambda_i)^p E_i = 0,$$

so that N is nilpotent. (Observe that $N = 0$ if $m_1 = \cdots = m_k = 1$.) We can write

$$T = \sum_{i=1}^{k} \lambda_i E_i + N. \tag{5.9–6}$$

This representation of T is a consequence of the fact that R_λ is rational. There is a converse to this, as the following theorem shows.

Theorem 5.9–F. *Suppose X is a complex Banach space. Let $E_1, \cdots, E_k$ be elements of $[X]$ which are projections such that $E_iE_j = 0$ if $i \neq j$, and $I = E_1 + \cdots + E_k$. Let T be defined by* (5.9–6), *where $\lambda_1, \cdots, \lambda_k$ are complex constants, and N is a nilpotent member of $[X]$ such that $NE_i = E_iN$ for each i. Then R_λ is rational.*

PROOF. Suppose $N^p = 0$. Now, $(T - \lambda_i)E_i = NE_i$, as we see from (5.9–6). Hence $(T - \lambda_i)^p E_i = N^p E_i = 0$. It now follows from $I = E_1 + \cdots + E_k$ that $(T - \lambda_1)^p \cdots (T - \lambda_k)^p = 0$. Therefore, by Theorem 5.9–B, R_λ is rational.

PROBLEMS

1. Let X be n-dimensional and suppose $T \in [X]$. Let $u_1, \cdots, u_n$ be a basis for X, and let $D(\lambda)$ be the determinant of the matrix corresponding to $\lambda - T$ when this basis is used. Then $D(T) = 0$. (This is the Hamilton-Cayley theorem.) Use Theorem 5.9–A and Cramer's rule. See problem 6 also.

2. Suppose that R_λ is rational, and adopt the notation of Theorem 5.9–E and Theorem 5.9–D*b*. Let N_i be the null manifold of $\lambda_i - T$, and let Y be the subspace generated by $N_1 \cup \cdots \cup N_k$. Then $Y = X$ if and only if $m_1 = \cdots = m_k = 1$.

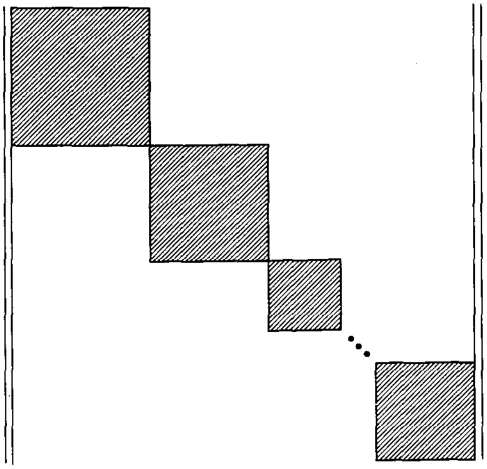
3. Suppose that R_λ is rational. Using the notation of Theorem 5.9–E and the paragraph which follows it, show that, if f is locally analytic on $\sigma(T)$,

$$f(T) = \sum_{i=1}^{k} \sum_{j=0}^{m_i-1} \frac{f^{(j)}(\lambda_i)}{j!}(T - \lambda_i)^j E_i.$$

In particular, suppose m is fixed, $1 \leqslant m \leqslant k$, and let $f(\lambda)$ be a polynomial such that $f^{(j)}(\lambda_i) = 1$ if $i = m$, $j = 0$, while $f^{(j)}(\lambda_i) = 0$ otherwise as it occurs in the formula for $f(T)$. Then $f(T) = E_m$. Since a polynomial of this kind exists (of degree at most $m_1 + \cdots + m_k - 1$), each E_m is expressible as a polynomial in T. See Hamburger and Grimshaw, **1**, page 111.

4. Suppose $T \in [X]$ and that $X_0 = \mathscr{R}(T)$ is finite dimensional, but $X_0 \neq X$. Then $0 \in \sigma(T)$. Observe that $T(X_0) \subset X_0$. Let T_0 be the restriction of T to X_0, and let S_λ be the resolvent of T_0. In order to see how to express R_λ in terms of S_λ, write $y = \lambda x - Tx$, $Ty = \lambda Tx - T_0(Tx)$, whence $Tx = S_\lambda Ty$ if $\lambda \notin \sigma(T_0)$. Then $y = \lambda x - S_\lambda Ty$, and $x = \lambda^{-1}[I + S_\lambda T]y$. This suggests that $R_\lambda = \lambda^{-1}[I + S_\lambda T]$ and that $\sigma(T) = (0) \cup \sigma(T_0)$. Verify that this is correct. Observe that R_λ is rational, since S_λ is. An important type of operator with a finite-dimensional range is an integral operator with "degenerate" kernel $k(s, t) = \alpha_1(s)\beta_1(t) + \cdots + \alpha_n(s)\beta_n(t)$.

5. If X is finite dimensional, formula (5.9–6) plus certain well-known facts about nilpotent operators provide the basis for obtaining the *Jordan normal form* matrix representation of the operator T. Choose a basis $u_1, \cdots, u_n$ for X in such a way that successive sets of u's form bases for $M_1, \cdots, M_k$. Then the matrix representation of T takes the form shown here in the adjacent diagram,



where the only nonzero elements are inside the shaded blocks. The i^{th} block, counting down the diagonal, is the matrix representation of the restriction of T to M_i. Owing to the nilpotency of N, whose restriction to M_i is the same as that of $T - \lambda_i$, the basis for M_i may be chosen in such a way that the matrix representation of the restriction of N to M_i has no nonzero elements except for the possibility of 1's in a certain arrangement along the diagonal directly below the main diagonal. There will be some such 1's if $m_i > 1$, but not if $m_i = 1$. Thus the matrix for T has

$$\underbrace{\lambda_1, \cdots, \lambda_1;}_{\nu_1} \quad \underbrace{\lambda_2, \cdots, \lambda_2;}_{\nu_2} \quad \cdots; \quad \underbrace{\lambda_k, \cdots, \lambda_k}_{\nu_k}$$

down the main diagonal, where ν_i is the dimension of M_i. The only other nonzero elements are the 1's already referred to, on the diagonal just below the main diagonal. For more details on the matrix representation of nilpotent operators see Halmos, **1**, pages 164–165.

6. Using the results in problem 5, show that the determinant $D(\lambda)$ in problem 1 is $(\lambda - \lambda_1)^{\nu_1} \cdots (\lambda - \lambda_k)^{\nu_k}$.

SUGGESTED READING

Bade, **1**.
Courant and Hilbert, **1**, parts of Chapter 3.
Dunford, **3**, **4**.
Dunford and Schwartz, **1**, Chapter 7.
Friedman, **1**, Chapter 2.
Halmos, **1**, Chapter 2.
Hamburger and Grimshaw, **1**, Chapter 4.
Hille, **1**, Chapters 5, 11, 15 and many parts of other chapters.
Lorch, **1**.
Ljusternik and Sobolew, **1**, Chapter 4.
Riesz, F., **1**, Chapter 4.
Riesz and Nagy, **1**, Chapters 4, 5, 11.
Stone, **1**, Chapter 4.
Taylor, **1**, **2**.
Titchmarsh, **2**.
Zaanen, **1**, Chapters 11, 13.

6

SPECTRAL ANALYSIS IN HILBERT SPACE

6.0 Introduction

The principal aim of this chapter is to study the more elementary properties of symmetric linear operators (also called Hermitian operators if the scalar field is complex) and especially the properties of self-adjoint and normal operators in Hilbert space. The theory of compact symmetric operators is developed without requiring the space to be complete. This material has applications to the Hilbert-Schmidt theory of integral equations with symmetric kernel and to symmetric differential operators with compact resolvent; in particular, to the classical Sturm-Liouville differential-equation problems. The spectral theorem is proved for bounded self-adjoint operators and for unitary operators. The state of affairs for unbounded self-adjoint operators is sketched without proofs.

The "big" theorems of the chapter are: 6.4–B, 6.4–C, and 6.4–D for compact symmetric operators; 6.5–A and 6.5–C for the spectral theorem for bounded self-adjoint operators and the operational calculus founded on it; and 6.5–B, the spectral theorem for unitary operators. Theorems 6.5–D and 6.5–E are of fundamental interest, since they relate the spectral classification of a point (in relation to a bounded self-adjoint operator) to the function-theoretic character of the resolution of the identity at that point.

Theorem 6.2–C, comparatively simple in itself, is of crucial importance in the proof of Theorem 6.5–A.

6.1 Bilinear and Quadratic Forms

Throughout this section X denotes a not necessarily complete inner-product space. The scalar field may be either real or complex, except when we specify one or the other explicitly. If no such explicit specification is made, a bar indicating the complex conjugate of a number is to be ignored if the situation under consideration is being regarded from the point of view of real scalars.

Suppose A is a linear operator with domain X and range in X. Then the inner product (Ax, y) is a linear functional of x for each fixed y. As a function of y, (Ax, y) is linear when the scalar field is real; in the complex case, however, we do not quite have linearity with respect to y, for $(Ax, \alpha y) = \bar{\alpha}(Ax, y)$, the scalar factor coming outside as $\bar{\alpha}$ rather than as α. However, for convenience in embracing both the real and complex cases in one terminology, we shall say that (Ax, y) is *bilinear* in x and y.

For some purposes it is desirable to study bilinearity directly, instead of through the medium of a linear operator. Hence we make the following definition: A scalar-valued function ϕ on $X \times X$ is called a *bilinear form* if $\phi(x, y)$ is linear in x for each y, while $\overline{\phi(x, y)}$ is linear in y for each x. With ϕ we associate the functional ψ on X defined by $\psi(x) = \phi(x, x)$. We call ψ the *quadratic form* corresponding to ϕ. Observe that $\psi(\alpha x) = |\alpha|^2\psi(x)$.

One important relation between ϕ and ψ is expressed by the formula

$$\tag{6.1–1} \tfrac{1}{2}\{\phi(x, y) + \phi(y, x)\} = \psi\left(\frac{x + y}{2}\right) - \psi\left(\frac{x - y}{2}\right).$$

It may be verified by expanding the terms on the right, using the bilinearity. When the scalar field is complex we have

$$\tag{6.1–2} \phi(x, y) = \psi\left(\frac{x + y}{2}\right) - \psi\left(\frac{x - y}{2}\right) + i\psi\left(\frac{x + iy}{2}\right) - i\psi\left(\frac{x - iy}{2}\right).$$

In the complex case, therefore, ϕ can be expressed entirely in terms of ψ. This is not always true in the real case, however, for in this case the left member of (6.1–1) is a bilinear form, in general not the same as ϕ, whose corresponding quadratic form is the same as the one corresponding to ϕ.

A bilinear form ϕ is continuous jointly in its two variables if and only if $|\phi(x, y)|$ is bounded for all x and y such that $\|x\| \leqslant 1$, $\|y\| \leqslant 1$ (proof similar to that of Theorem 3.1–A). When ϕ is continuous we define

$$\|\phi\| = \sup \frac{|\phi(x, y)|}{\|x\|\,\|y\|} \qquad (\text{for } x \neq 0, y \neq 0),$$

or, equivalently, we may take the supremum merely for all x and y with $\|x\| = \|y\| = 1$. To avoid trivialities we are assuming that the space X does not reduce to (0). For the corresponding quadratic form we define

$$\|\psi\| = \sup_{\|x\|=1} |\psi(x)| = \sup_{x\neq 0} \frac{|\psi(x)|}{\|x\|^2}.$$

Obviously $\|\psi\| \leqslant \|\phi\|$. When the field of scalars is complex it is easily seen with the aid of (6.1–2) and (4.81–1) that $\|\phi\| \leqslant 2\|\psi\|$. Hence

$$\|\psi\| \leqslant \|\phi\| \leqslant 2\|\psi\| \tag{6.1–3}$$

in the complex case.

A bilinear form ϕ is called *symmetric* if $\phi(x, y) = \overline{\phi(y, x)}$ (in the complex case such a form is also called *Hermitian*). If ϕ is symmetric we can prove that

$$\|\phi\| = \|\psi\|. \tag{6.1–4}$$

We have only to prove that $\|\phi\| \leqslant \|\psi\|$. Owing to the symmetry we have from (6.1–1) and (4.81–1):

$$|\operatorname{Re} \phi(x, y)| \leqslant \tfrac{1}{2}\|\psi\| (\|x\|^2 + \|y\|^2)$$

(with the Re symbol, for the "real part," superfluous in the real case). For fixed x and y with $\|x\| = \|y\| = 1$ we can choose α so that $|\alpha| = 1$ and $\alpha\phi(x, y) = |\phi(x, y)|$, whence

$$|\phi(x, y)| = \phi(\alpha x, y) = |\operatorname{Re} \phi(\alpha x, y)| \leqslant \|\psi\|,$$

and so $\|\phi\| \leqslant \|\psi\|$.

Theorem 6.1–A. *Suppose $\phi(x, y) = (Ax, y)$, where A is linear on X into X. Then ϕ is continuous if and only if A is continuous, and then $\|A\| = \|\phi\|$.*

PROOF. If A is continuous, $|(Ax, y)| \leqslant \|Ax\|\,\|y\| \leqslant \|A\|\,\|x\|\,\|y\|$. If ϕ is continuous, $\|Ax\|^2 = (Ax, Ax) \leqslant \|\phi\|\,\|x\|\,\|Ax\|$, or $\|Ax\| \leqslant \|\phi\|\,\|x\|$. The conclusions now follow.

Theorem 6.1–B. *Suppose that X is complete and that ϕ is a continuous bilinear form. Then there exists $A \in [X]$ such that $\phi(x, y) = (Ax, y)$.*

PROOF. For fixed x, $\overline{\phi(x, y)}$ is a continuous linear functional of y. Hence, by Theorem 4.81–C, we can represent the functional in the form

$\overline{\phi(x, y)} = (y, Ax)$, where Ax is some vector depending on x. This defines an operator A on X; the linearity of A is easily verified, since $(Ax, y) = \phi(x, y)$. That A is continuous follows from Theorem 6.1–A.

PROBLEM

Suppose A is linear on X into X, where X is a complex inner-product space. Then, if $(Ax, x) = 0$ for each $x \in X$, it follows that $A = 0$. Is this true for real inner-product spaces?

6.11 Symmetric Operators

A linear operator A with domain and range in the inner-product space X is called *symmetric* if

$$(Ax, y) = (x, Ay) \tag{6.11–1}$$

for each x and y in $\mathscr{D}(A)$. In case X is a complex space, a symmetric linear operator is also called a *Hermitian* operator. If we regard $\mathscr{D}(A)$ by itself as an inner-product space, $\phi(x, y) = (Ax, y)$ is a bilinear form on $\mathscr{D}(A) \times \mathscr{D}(A)$, and ϕ is symmetric (as defined in § 6.1) if and only if A is symmetric. The corresponding quadratic form, defined on $\mathscr{D}(A)$, is (Ax, x).

We assume that A is symmetric, with $\mathscr{D}(A) \neq (0)$. Then (Ax, x) is real, even when X is a complex space, and we define

$$m(A) = \inf_{\|x\|=1} (Ax, x), \qquad M(A) = \sup_{\|x\|=1} (Ax, x). \tag{6.11–2}$$

The possibilities $m(A) = -\infty$, $M(A) = +\infty$ are not excluded.

Theorem 6.11–A. *If A is symmetric and λ is an eigenvalue of A, then λ is real, and $m(A) \leqslant \lambda \leqslant M(A)$. Eigenvectors corresponding to distinct eigenvalues are orthogonal.*

PROOF. Suppose $\|x\| = 1$ and $Ax = \lambda x$. Then $(Ax, x) = (\lambda x, x) = \lambda$, so λ is real and $m(A) \leqslant \lambda \leqslant M(A)$. If $Ax = \lambda x$ and $Ay = \mu y$, where $\lambda \neq \mu$, we have $\lambda(x, y) = (\lambda x, y) = (Ax, y) = (x, Ay) = (x, \mu y) = \mu(x, y)$, or $(\lambda - \mu)(x, y) = 0$, whence $(x, y) = 0$.

Next we consider the question as to whether the bounds $m(A)$ and $M(A)$ can be attained by values of (Ax, x) when $\|x\| = 1$. If $\lambda = m(A)$ happens to be an eigenvalue of A, there is an x with $\|x\| = 1$ and $Ax = \lambda x$. Then $(Ax, x) = (\lambda x, x) = m(A)$. A similar thing is true if $M(A)$ is an eigenvalue. There is a valid converse if $\mathscr{D}(A)$ is dense in X.

Theorem 6.11–B. *Suppose that A is symmetric and that $\mathscr{D}(A)$ is dense in X. Suppose $y \in \mathscr{D}(A)$, $\|y\| = 1$, and $(Ay, y) = \lambda$, where λ is either $m(A)$ or $M(A)$. Then $Ay = \lambda y$, so that λ is an eigenvalue of A.*

PROOF. Suppose for example that $\lambda = m(A)$. Let $B = A - \lambda$. Then B is symmetric, with $(By, y) = 0$ and $m(B) = 0$, so that $(Bx, x) \geqslant 0$ for every x in $\mathscr{D}(A)$. Let $\alpha = \theta(By, x)$, where x is arbitrary in $\mathscr{D}(A)$ and θ is an arbitrary real number. Then

$$0 \leqslant (B[y + \alpha x], y + \alpha x) = (By, y) + \alpha(Bx, y) + \bar{\alpha}(By, x) + |\alpha|^2(Bx, x),$$

or

$$0 \leqslant \theta|(x, By)|^2\{2 + \theta(Bx, x)\}.$$

Since the right member of the inequality changes sign with θ when θ is small, unless $(x, By) = 0$, we conclude that $(x, By) = 0$. This is true for each $x \in \mathscr{D}(A)$, and, since $\mathscr{D}(A)$ is dense in X, we conclude that $By = 0$ or $Ay = \lambda y$. If $\lambda = M(A)$, we put $B = \lambda - A$, and the same argument applies.

When X is complete and A is symmetric, with $\mathscr{D}(A) = X$, we can prove (see § 6.2) that A is continuous and that $\sigma(A)$ lies on the interval $m(A) \leqslant \lambda \leqslant M(A)$ of the real axis, with $m(A)$ and $M(A)$ actually belonging to $\sigma(A)$.

In the next theorem we consider the situation in which $\mathscr{D}(A) = X$, though X need not be complete.

Theorem 6.11–C. *If A is symmetric, with $\mathscr{D}(A) = X$, then A is continuous if and only if $m(A)$ and $M(A)$ are both finite, and in that case*

$$\|A\| = \max\{|m(A)|, |M(A)|\}. \tag{6.11–3}$$

PROOF. We know from Theorem 6.1–A that A is continuous if and only if $|(Ax, y)|$ is bounded for all x and y for which $\|x\| = \|y\| = 1$; and, as we see in conjunction with (6.1–4), this is the same as demanding that $\sup\limits_{\|x\|=1} |(Ax, x)|$ be finite, this supremum then being equal to $\|A\|$. Formula (6.11–3) now follows from the definitions (6.11–2).

Example 1. Let X be the Hilbert space $L^2(0, 2\pi)$. Let $\mathscr{D}(A)$ be the subspace of X determined by those functions $x(t)$ which are absolutely continuous on $[0, 2\pi]$, such that $x(0) = x(2\pi)$ and such that the derivative $x'(t)$ is of class $\mathscr{L}^2(0, 2\pi)$. For $x \in \mathscr{D}(A)$ define $Ax = y$ to mean $y(t) = -ix'(t)$. Then A is symmetric. For, if $x, y \in \mathscr{D}(A)$,

$$i\frac{d}{dt}\{x(t)\overline{y(t)}\} = ix(t)\overline{y'(t)} + ix'(t)\overline{y(t)},$$

and so

$$0 = ix(t)\overline{y(t)}\Big|_0^{2\pi} = (x, Ay) - (Ax, y).$$

Example 2. An important type of symmetric operator arises in connection with certain boundary-value problems for second-order ordinary differential equations. The symmetry of the operator in such problems depends both on the nature of the boundary conditions and on certain formal properties of the differential operator which occurs. The space X is taken to be either $L^2(a, b)$ or some subspace of $L^2(a, b)$, with the usual inner product of $L^2(a, b)$. When the domain $\mathscr{D}(A)$ is suitably defined, $Ax = u$ is expressed by

$$u(t) = -\frac{d}{dt}[p(t)x'(t)] + q(t)x(t) \tag{6.11–4}$$

where p and q are certain *real-valued* functions. This formula exhibits the essential formal structure of the differential operator. There must be conditions on p, q and $\mathscr{D}(A)$ which insure that Ax is a well-defined member of X when $x \in \mathscr{D}(A)$. For instance, if (a, b) is a finite interval and if X consists of those elements of $L^2(a, b)$ which correspond to functions continuous on $[a, b]$, we might require q to be continuous and p to be continuously differentiable. We could then take $\mathscr{D}(A)$ to consist of those functions which are twice continuously differentiable and satisfy certain boundary conditions. If X is taken to be all of $L^2(a, b)$, we might require that q be measurable and bounded and that p be absolutely continuous. Then $\mathscr{D}(A)$ could be taken to consist of those elements of $L^2(a, b)$ corresponding to functions $x(t)$ which are such that both $x(t)$ and $p(t)x'(t)$ are absolutely continuous, $p(t)x'(t)$ has a derivative of class $\mathscr{L}^2(a, b)$, and $x(t)$ satisfies certain boundary conditions.

If $Ax = u$ and $Ay = v$, it is easily verified that we have

$$u(t)\overline{y(t)} - x(t)\overline{v(t)} = \frac{d}{dt}\{p(t)[x(t)\overline{y'(t)} - x'(t)\overline{y(t)}]\}.$$

Thus

$$(Ax, y) - (x, Ay) = p(t)[x(t)\overline{y'(t)} - x'(t)\overline{y(t)}]\Big|_a^b.$$

The operator A is then seen to be symmetric, provided that the definition of $\mathscr{D}(A)$ insures that

$$p(b)[x(b)\overline{y'(b)} - x'(b)\overline{y(b)}] = p(a)[x(a)\overline{y'(a)} - x'(a)\overline{y(a)}]. \tag{6.11–5}$$

In case the interval is infinite, this condition must be interpreted appropriately. For a finite interval the boundary conditions are of the form

$$\alpha_{11}x(a) + \alpha_{12}x'(a) + \beta_{11}x(b) + \beta_{12}x'(b) = 0$$
$$\alpha_{21}x(a) + \alpha_{22}x'(a) + \beta_{21}x(b) + \beta_{22}x'(b) = 0$$

where the α's and β's are certain real scalars. For these boundary conditions it can be shown that condition (6.11–5) is equivalent to the condition

$$p(a)\begin{vmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{vmatrix} = p(b)\begin{vmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{vmatrix}. \tag{6.11–6}$$

It is not our intent in this book to discuss in detail the spectral analysis of a symmetric operator A of the type just described. However, in § 6.4, as an application of the theory of compact symmetric operators, we do discuss the solution of boundary-value problems of this type in the special case in which A^{-1} is a compact operator belonging to $[X]$.

We return now to the general consideration of symmetric operators. Suppose A is symmetric, and that $(Ax, x) \geqslant 0$ for each $x \in \mathscr{D}(A)$. Let us write $\{x, y\} = (Ax, y)$. Then $\{x, y\}$, as a function on $\mathscr{D}(A) \times \mathscr{D}(A)$, has all the properties of an inner product except possibly the property that $\{x, x\} = 0$ implies $x = 0$. This last property is not needed in the derivation of the Schwarz inequality $|\{x, y\}|^2 \leqslant \{x, x\}\{y, y\}$, however (see Theorem 3.2–A), and so we obtain the inequality

$$|(Ax, y)|^2 \leqslant (Ax, x)(Ay, y), \qquad x, y \in \mathscr{D}(A) \tag{6.11–7}$$

under the stated conditions on A. This result is used in the proof of Theorem 6.2–B.

PROBLEMS

1. Let A be linear on X into X, where X is a complex inner-product space. Suppose (Ax, x) is real for every x. Then A is symmetric. Use (6.1–2).

2. Suppose A is linear on X into X and symmetric. Then, if $(Ax, x) = 0$ for each x, it follows that $A = 0$. Compare with the problem at the end of § 6.1.

6.12 A Theorem of Schur

A bounded linear operator A with domain l^2 and range in l^2 is representable by an infinite matrix (α_{ij}); see Example 5, § 4.2 and problems 3 and 4, § 4.51. There is a useful theorem, due to I. Schur, **1**, which gives conditions on an infinite matrix sufficient to insure that it defines an element A of $[l^2]$. Schur's theorem also gives an estimate of $\|A\|$. The proof makes use of some facts about symmetric operators in finite-dimensional spaces.

Theorem 6.12–A. *Let* (α_{ij}) *be an infinite matrix of scalars such that*

$$\sum_{j=1}^{\infty} |\alpha_{ij}| \leqslant M_1, \qquad i = 1, 2, \cdots \tag{6.12–1}$$

and

$$\sum_{i=1}^{\infty} |\alpha_{ij}| \leqslant M_2, \qquad j = 1, 2, \cdots. \tag{6.12–2}$$

Then the matrix represents an operator A *on* l^2 *to* l^2 *such that*

$$\|A\| \leqslant (M_1 M_2)^{\frac{1}{2}}. \tag{6.12–3}$$

PROOF. From (6.12–2) it is clear that $\sum_k |\alpha_{kj}|^2 < \infty$ for each j. Hence the series

$$\beta_{ij} = \sum_{k=1}^{\infty} \alpha_{ki}\bar{\alpha}_{kj}$$

is convergent. Now consider the linear operator B_n on $l^2(n)$ defined by

$$\eta_i = \sum_{j=1}^{n} \beta_{ij}\xi_j, \qquad i = 1, \cdots, n.$$

That is, B_n maps $x = (\xi_1, \cdots, \xi_n)$ into $y = (\eta_1, \cdots, \eta_n)$. The inner product of $B_n x$ and x is

$$(B_n x, x) = \sum_{i=1}^{n}\sum_{j=1}^{n} \beta_{ij}\xi_j\bar{\xi}_i = \sum_{i}\sum_{j}\sum_{k=1}^{\infty} \alpha_{ki}\bar{\alpha}_{kj}\xi_j\bar{\xi}_i,$$

or

$$(B_n x, x) = \sum_{k=1}^{\infty}\left|\sum_{i=1}^{n} \alpha_{ki}\bar{\xi}_i\right|^2. \tag{6.12–4}$$

Observe that B_n is symmetric, because $\beta_{ij} = \bar{\beta}_{ji}$. The eigenvalues of B_n are nonnegative, since $(B_n x, x) \geqslant 0$. Suppose λ is an eigenvalue, say $B_n x = \lambda x$, where $x = (\xi_1, \cdots, \xi_n) \neq 0$. Let $|\xi_p|$ be the largest of the values $|\xi_1|, \cdots, |\xi_n|$. Then

$$\lambda\xi_p = \sum_{j=1}^{n} \beta_{pj}\xi_j, \qquad \lambda|\xi_p| \leqslant \sum_{j=1}^{n} |\beta_{pj}|\,|\xi_p|,$$

and so

$$\lambda \leqslant \sum_{j=1}^{n} |\beta_{pj}| = \sum_{j=1}^{n}\left|\sum_{k=1}^{\infty} \alpha_{kp}\bar{\alpha}_{kj}\right|,$$

$$\lambda \leqslant \sum_{k=1}^{\infty} |\alpha_{kp}| \sum_{j=1}^{n} |\alpha_{kj}| \leqslant M_1 M_2.$$

Now, the largest eigenvalue of B_n is given by

$$\lambda = \max_{\|x\|=1} (B_n x, x);$$

we know this from Theorems 6.11–A, 6.11–B. We have thus proved that $(B_n x, x) \leqslant M_1 M_2$ if $\|x\| = 1$, which implies that

$$(B_n x, x) \leqslant M_1 M_2 \sum_{j=1}^{n} |\xi_j|^2$$

for arbitrary $(\xi_1, \cdots, \xi_n)$. In view of (6.12–4) we infer that

$$\sum_{k=1}^{\infty} \left| \sum_{i=1}^{\infty} \alpha_{ki} \xi_i \right|^2 \leqslant M_1 M_2 \sum_{j=1}^{\infty} |\xi_j|^2$$

for each vector $x = \{\xi_j\}$ in l^2. This proves (6.12–3).

The inequality (6.12–3) is a special case of the result of problem 1, § 4.52.

6.2 Normal and Self-adjoint Operators

Throughout this section X denotes a complete inner-product space, and all of the operators which we consider in the theorems belong to $[X]$. We assume $X \neq (0)$. If $A \in [X]$, then $A^* \in [X]$ also (see § 4.9); the definition of A^* is fully expressed by the relation

$$(6.2\text{–}1) \qquad (Ax, y) = (x, A^*y) \qquad x, y \in X.$$

We recall from § 4.9 that A is called *normal* if $AA^* = A^*A$, and *self-adjoint* if $A = A^*$. We see that a self-adjoint operator is symmetric. Conversely, a symmetric operator whose domain is all of X (with X complete) must be self-adjoint. The proof of this fact hinges upon the closed-graph theorem, or alternatively, upon the principle of uniform boundedness. See problem 1 at the end of this section. For a related matter in a more general setting see problem 1, § 4.5.

There is a generalization of the definition of self-adjointness, so that certain unbounded operators with domains dense in X are called self-adjoint. These operators are symmetric, but not all symmetric operators are self-adjoint in this extended sense. See § 6.7.

If $X = L^2(a, b)$ and if K is an operator defined by an $\mathscr{L}^2$-kernel $k(s, t)$, the adjoint K^* is defined by the kernel $\overline{k(t, s)}$. In this case, therefore, K is self-adjoint if and only if $k(s, t) =^0 \overline{k(t, s)}$.

If $X = l^2$ and if $A \in [l^2]$ is represented by the infinite matrix (α_{ij}), the adjoint is represented by $(\bar{\alpha}_{ji})$ (i.e., the conjugate of the transposed matrix). Hence $A = A^*$ means that $\alpha_{ij} = \bar{\alpha}_{ji}$ for each i and j.

Our first theorem concerns the localization of the spectrum of any member of $[X]$.

Definition. If $A \in [X]$, let $V(A)$ be the closure of the set of values of (Ax, x) for all x with $\|x\| = 1$.

Theorem 6.2–A. *If $A \in [X]$, $\sigma(A) \subset V(A)$. If the distance d from λ to $V(A)$ is positive, then*

$$\|(\lambda - A)^{-1}\| \leqslant d^{-1}. \tag{6.2–1}$$

PROOF. Suppose λ is not in $V(A)$, the distance from λ to $V(A)$ being $d > 0$. Then $|(Ax, x) - \lambda| \geqslant d$ if $\|x\| = 1$. We can write this as $|(Ax, x) - \lambda(x, x)| = |(Ax - \lambda x, x)| \geqslant d$ if $\|x\| = 1$, whence $|(Ax - \lambda x, x)| \geqslant d\|x\|^2$ for every x. But then $d\|x\|^2 \leqslant \|Ax - \lambda x\|\,\|x\|$, or $d\|x\| \leqslant \|Ax - \lambda x\|$. This shows that $(\lambda - A)^{-1}$ exists and is continuous, with

$$\|(\lambda - A)^{-1}y\| \leqslant d^{-1}\|y\| \tag{6.2–2}$$

for every y in the range of $\lambda - A$. We know, then, that λ is either in $\rho(A)$ or $R\sigma(A)$. It remains only to show that λ cannot be in $R\sigma(A)$. Now, if $\lambda \in R\sigma(A)$, the first of the relations (4.9–6) shows that the null space of $(\lambda - A)^* = \bar{\lambda} - A^*$ contains a nonzero vector, i.e., that $\bar{\lambda}$ is an eigenvalue of A^*. Suppose $\|x\| = 1$ and $A^*x = \bar{\lambda}x$. Then $(Ax, x) = (x, A^*x) = (x, \bar{\lambda}x) = \lambda$, so that $\lambda \in V(A)$. This contradiction finishes the proof.

It can be shown that $V(A)$ is a convex set (see Stone, **1**, pages 130–133). In the case when A is normal, $V(A)$ is the smallest closed convex set containing $\sigma(A)$ (see A. Wintner, **2**, page 248, and Stone, **1**, pages 327–328). For finite-dimensional spaces these results go back to Toeplitz and Hausdorff; see Wintner, **1**, pages 33–38.

Theorem 6.2–B. *Suppose A is a bounded self-adjoint operator. Then $\sigma(A)$ lies on the closed interval $[m(A), M(A)]$ of the real axis. The end points of this interval belong to $\sigma(A)$.*

PROOF. The first assertion is a consequence of Theorem 6.2–A, for it follows from (6.11–2) that $V(A)$ lies on the closed interval in question. (In fact, $V(A)$ *is* the closed interval $[m(A), M(A)]$, as a result of the fact that $V(A)$ is convex; but we do not need this information.) To see that $m(A)$ is in $\sigma(A)$ let $\lambda = m(A)$ and observe that $([A - \lambda]x, x) \geqslant 0$ for each x. Therefore, by (6.11–7), with $A - \lambda$, x, and $(A - \lambda)x$ in place of A, x, and y respectively, we have

$$\begin{aligned} \|(A - \lambda)x\|^4 &\leqslant ([A - \lambda]x, x)([A - \lambda]^2x, [A - \lambda]x) \\ &\leqslant ([A - \lambda]x, x)\,\|A - \lambda\|^3\|x\|^2. \end{aligned}$$

If $\|A - \lambda\| = 0$ it is clear that $\lambda \in \sigma(A)$. Otherwise we see from the foregoing that

$$\inf_{\|x\|=1} \|(A - \lambda)x\| = 0,$$

whence $\lambda \in \sigma(A)$. This is because $\inf_{\|x\|=1} ([A - \lambda]x, x) = 0$, by the definition of $m(A)$. The proof that $M(A)$ is in $\sigma(A)$ is similar.

If λ is not on the interval $[m(A), M(A)]$ we can estimate the norm of $(\lambda - A)^{-1}$ in terms of the distance from λ to the interval, using (6.2–1). In particular, if X is a complex space and λ is not real, the distance in question is not less than the absolute value of the imaginary part of λ. Hence

$$\|(\lambda - A)^{-1}\| \leqslant 2/|\lambda - \bar{\lambda}|. \tag{6.2–3}$$

Theorem 6.2–C. *If $A \in [X]$ and $A = A^*$, the spectral radius of A is $\|A\|$.*

PROOF. This is an immediate consequence of (6.11–3) and Theorem 6.2–B.

This result, which is of crucial importance in our proof of the fundamental spectral theorem for self-adjoint operators (Theorem 6.5–A), can be proved in other ways (see problem 2). The property is true, more generally, of normal operators in complex spaces, as we shall see presently.

Theorem 6.2–D. *If $A \in [X]$, A is normal if and only if $\|Ax\| = \|A^*x\|$ for every x.*

PROOF. We have

$$\|Ax\|^2 = (Ax, Ax) = (x, A^*Ax),$$

and

$$\|A^*x\|^2 = (A^*x, A^*x) = (x, A^{**}A^*x) = (x, AA^*x).$$

Thus $\|Ax\| = \|A^*x\|$ if and only if $(x, A^*Ax) = (x, AA^*x)$. But this is true for every x if and only if $A^*A = AA^*$. For, we can set $B = A^*A - AA^*$, and then $B^* = B$. The desired result now follows by noting that $(Bx, x) = 0$ for all x is equivalent to $B = 0$. See § 6.11, problem 2.

Theorem 6.2–E. *If the space is complex and A is normal, the spectral radius of A is exactly $\|A\|$.*

PROOF. We first prove that $\|A^2\| = \|A\|^2$. We have $\|A^2x\| = \|A^*Ax\|$ for every x, by Theorem 6.2–D. Thus $\|A^2\| = \|A^*A\|$. But we know that $\|A^*A\| = \|A\|^2$ (by Theorem 4.9–B). Since powers of A are normal, it now follows by induction that $\|A^p\| = \|A\|^p$ if p is a positive integer of the form 2^n. The fact that the spectral radius of A is $\|A\|$ now follows from Theorem 5.2–E.

Next we consider the ascent of a normal operator.

Theorem 6.2–F. *If A is normal, its ascent is either* 0 *or* 1.

PROOF. We give the proof first on the assumption that $A = A^*$. If $x \in \mathcal{N}(A^2)$, we have $0 = (A^2x, x) = (Ax, Ax)$, whence $Ax = 0$. We conclude that $\mathcal{N}(A^2) \subset \mathcal{N}(A)$, whence $\mathcal{N}(A^2) = \mathcal{N}(A)$, and the ascent of A does not exceed 1. For the case of an arbitrary normal A, let $B = AA^* = A^*A$, so that $B^* = B$. If $x \in \mathcal{N}(A^2)$, then $B^2x = (AA^*)^2x = (A^*)^2A^2x = 0$. Hence $Bx = 0$, by the first part of the proof. Then $0 = (A^*Ax, x) = (Ax, Ax)$, or $Ax = 0$, and so $\mathcal{N}(A^2) = \mathcal{N}(A)$, as before. This finishes the proof.

We also have:

Theorem 6.2–G. *If A is normal, $\overline{\mathscr{R}(A)}$ and $\mathcal{N}(A)$ are orthogonal complements, so that $X = \overline{\mathscr{R}(A)} \oplus \mathcal{N}(A)$.*

PROOF. We know from Theorem 6.2–D that $\mathcal{N}(A) = \mathcal{N}(A^*)$. We then see from (4.9–6) that $\overline{\mathscr{R}(A)} = \mathcal{N}(A)^{\perp}$, and the result follows from Theorem 4.82–A.

Now suppose that X is a complex space. If A is normal, so is $\lambda - A$. Let R_λ be the resolvent of A when $\lambda \in \rho(A)$. If λ_0 is a pole of R_λ, it follows by Theorems 5.8–A, 6.2–F, and 6.2–G that the pole is of first order, that $\mathscr{R}(\lambda_0 - A)$ is closed and that $\mathscr{R}(\lambda_0 - A)$ and $\mathcal{N}(\lambda_0 - A)$ are orthogonal complements. In this case, of course, λ_0 is an eigenvalue of A. Conversely, suppose that A is normal and that $\mathscr{R}(\lambda_0 - A)$ is a closed and proper subspace of X. Then, by Theorem 5.8–D, λ_0 is a first order pole of the resolvent of A.

If X is complex and finite dimensional and if A is normal, with the distinct eigenvalues $\lambda_1, \cdots, \lambda_k$, the ascent of $\lambda_i - A$ is 1 for each i (by Theorem 6.2–F). If M_i is the null manifold of $\lambda_i - A$, it follows that $X = M_1 \oplus \cdots \oplus M_k$ and that $A = \lambda_1E_1 + \cdots + \lambda_kE_k$, where E_i is the projection of X onto M_i determined by this particular direct sum decomposition of X. The reader is referred to Theorem 5.9–E and formula (5.9–6); in this case $N = 0$. The null manifolds $M_1, \cdots, M_k$ are mutually orthogonal (see problem 8). We have here a special case of the spectral theorem for normal operators.

PROBLEMS

1. Suppose A and B are linear on X into X and that $(Ax, y) = (x, By)$ for all x and y. Then $A \in [X]$ and $B = A^*$. One method of proof starts by showing that A is closed, whence $A \in [X]$. Another method: $|(Ax, y| = |(x, By)| \leqslant$

$\|x\| \, \|By\| \leqslant \|By\|$ for each y if $\|x\| \leqslant 1$. Then A is continuous, by the principle of uniform boundedness.

2. If $A \in [X]$ and $A = A^*$, prove that $r_\sigma(A) = \|A\|$, starting as follows: Let $\alpha = \|A\|$. Show that

$$\|(\alpha^2 - A^2)x\|^2 = \alpha^4\|x\|^2 - 2\alpha^2\|Ax\|^2 + \|A^2x\|^2.$$

Choose x_n so that $\|x_n\| = 1$ and $\|Ax_n\| \to \|A\|$, and deduce that $\alpha^2 \in \sigma(A^2)$. Explain why either $\alpha \in \sigma(A)$ or $-\alpha \in \sigma(A)$, and finish the proof.

3. If A is normal, so is $\lambda - A$.

4. Suppose A and B are normal and that $AB^* = B^*A$. Then $A^*B = BA^*$, and AB, BA, and $A + B$ are all normal.

5. If $A \in [X]$, we can write A in the form $A = H + iK$, where H and K are self-adjoint, in one and only one way, namely with $H = {}^1\!/_2(A + A^*)$, $K = (1/2i)(A - A^*)$. Then A is normal if and only if $HK = KH$.

6. Suppose $A \in [X]$. Then (*a*) $\lambda \in \rho(A)$ if and only if $\bar{\lambda} \in \rho(A^*)$; (*b*) $\lambda \in C\sigma(A)$ implies $\bar{\lambda} \in C\sigma(A^*)$; (*c*) $\lambda \in R\sigma(A)$ implies $\bar{\lambda} \in P\sigma(A^*)$; (*d*) $\lambda \in P\sigma(A)$ implies $\bar{\lambda} \in P\sigma(A^*) \cup R\sigma(A^*)$.

7. Suppose A is normal. Then λ is in the resolvent set, the continuous spectrum, or the point spectrum of A if and only if $\bar{\lambda}$ is in the corresponding set associated with A^*. The residual spectrum of A is empty. Moreover, λ is an eigenvalue of A if and only if the range of $\lambda - A$ is not dense in X. This follows from Theorem 4.9–C, applied to $\lambda - A$. A proof without recourse to the state diagram can be based on the three following facts (for an arbitrary λ): (*a*) $\mathscr{R}(\lambda - A)$ and $\mathscr{R}(\bar{\lambda} - A^*)$ have the same closure; (*b*) $\|(\lambda - A)x\| = \|(\bar{\lambda} - A^*)x\|$; (*c*) $\overline{\mathscr{R}(\lambda - A)}^{\perp} = \mathscr{N}(\bar{\lambda} - A^*)$.

8. Suppose A is normal. If λ is an eigenvalue of A and M is the corresponding eigenmanifold, M is also the eigenmanifold corresponding to $\bar{\lambda}$ as an eigenvalue of A^*. Eigenvectors corresponding to distinct eigenvalues of A are orthogonal.

9. Suppose A is normal and $A = H + iK$, where H and K are self-adjoint and $HK = KH$ (see problem 5). Then A^{-1} exists and belongs to $[X]$ if and only if $(H^2 + K^2)^{-1}$ exists and belongs to $[X]$. In that case $A^{-1} = A^*(H^2 + K^2)^{-1}$.

10. If A is normal, $\overline{\mathscr{R}(A^2)} = \overline{\mathscr{R}(A)}$.

11. If A is normal and its descent is finite, the range of A is closed [use Theorem 5.41–G and (5.41–4)].

6.3 Orthogonal Projections

Suppose that X is an inner-product space and that P is a projection defined on X. We know that $X = \mathscr{R}(P) \oplus \mathscr{N}(P)$ (see § 4.8). If $\mathscr{R}(P)$ and $\mathscr{N}(P)$ are orthogonal, we say that P is an *orthogonal projection.*

Theorem 6.3–A. *A projection is orthogonal if and only if it is symmetric.*

PROOF. Write $x = Px + u$, $y = Py + v$, with u and v in $\mathscr{N}(P)$. Then

$$(Px, y) = (Px, Py) + (Px, v),$$

$$(x, Py) = (Px, Py) + (u, Py).$$

If $\mathscr{R}(P)$ and $\mathscr{N}(P)$ are orthogonal, we see that $(Px, y) = (Px, Py) = (x, Py)$, so that P is symmetric.

On the other hand, if P is symmetric and $x \in \mathscr{R}(P)$, $y \in \mathscr{N}(P)$, then $Px = x$, $Py = 0$, so that $(x, y) = (Px, y) = (x, Py) = (x, 0) = 0$, so that P is orthogonal.

Theorem 6.3–B. *An orthogonal projection is continuous, and* $0 \leqslant (Px, x) \leqslant 1$ *if* $\|x\| = 1$. *If* $0 \neq P$, *then* $\|P\| = 1$; *if also* $P \neq I$, *then* $m(P) = 0$ *and* $M(P) = 1$.

PROOF. By the orthogonality of P, $x = Px + u$, with $(u, Px) = 0$. Then $\|x\|^2 = \|Px\|^2 + \|u\|^2$, whence $\|Px\|^2 \leqslant \|x\|^2$ and $\|P\| \leqslant 1$. From $P = P^2$ we see that $\|P\| \leqslant \|P\|^2$, whence $1 \leqslant \|P\|$ and so $\|P\| = 1$ if $P \neq 0$. If $\|x\| = 1$, we have $(Px, x) = (P^2x, x) = (Px, Px) \leqslant \|P\|^2 \leqslant 1$. Hence $0 \leqslant m(P)$ and $M(P) \leqslant 1$ [see (6.11–2)]. If $0 \neq P$, $Px = x$ can occur with $x \neq 0$, so $M(P) = 1$; if $P \neq I$, $Px = 0$ can occur with $x \neq 0$, so $m(P) = 0$.

For symmetric linear operators defined on X it is useful to introduce a partial-order relation. If A and B are symmetric and $(Ax, x) \leqslant (Bx, x)$ for every x, we write $A \leqslant B$. This is equivalent to $m(B - A) \geqslant 0$. If $A \leqslant B$, then $\alpha A \leqslant \alpha B$ when $\alpha > 0$, and $A + C \leqslant B + C$ for every symmetric C. If $A \leqslant B$ and $B \leqslant A$, then $A = B$, by problem 2, § 6.11.

Observe, by Theorem 6.3–B, that $0 \leqslant P \leqslant I$ if P is a symmetric projection.

Theorem 6.3–C. *If* P_1 *and* P_2 *are symmetric projections,* $P_1 \leqslant P_2$ *is equivalent to* $P_2P_1 = P_1$. *In this case* $P_2P_1 = P_1P_2$, *and* $P_2 - P_1$ *is also a symmetric projection.*

PROOF. From $P_2P_1 = P_1$ we have $(P_1P_2x, y) = (P_2x, P_1y) = (x, P_2P_1y) = (x, P_1y) = (P_1x, y)$, whence $P_1P_2 = P_1$. It is now easy to verify that $P_2 - P_1$ is a symmetric projection, whence $0 \leqslant P_2 - P_1$, or $P_1 \leqslant P_2$. Conversely, suppose $P_1 \leqslant P_2$. Then $I - P_2 \leqslant I - P_1$. Let $Q_k = I - P_k$. Then $Q_2 = Q_2{}^2$, $Q_2 \leqslant Q_1$, $Q_1P_1 = 0$, and so $(Q_2P_1x, Q_2P_1x) = (Q_2P_1x, P_1x) \leqslant (Q_1P_1x, P_1x) = 0$, whence $Q_2P_1 = 0$. This is the same as $P_1 - P_2P_1 = 0$, or $P_2P_1 = P_1$.

The foregoing theorem is used in the proof of Theorem 6.5–A.

6.4 Compact Symmetric Operators

For this section we assume that X is a real or complex inner-product space, with $X \neq (0)$. Completeness is assumed only as needed. We consider an operator $A \in [X]$, assuming that it is compact and symmetric.

To avoid trivialities we assume $A \neq 0$. For the definition of a compact operator see § 5.5. The present section is largely independent of § 5.5.

The basic fact about the type of operator here considered is that it possesses a finite or countably infinite set of nonzero eigenvalues (and *at least one* such eigenvalue); moreover, the structure of the operator can be completely analyzed in terms of the eigenmanifolds corresponding to these eigenvalues. The exact story is told by Theorem 6.4–B. What we have here is a generalization of the fact that, if X is finite dimensional and A is a symmetric member of $[X]$, a basis consisting of eigenvectors can be chosen for X in such a way that the matrix representing A is a diagonal matrix with each diagonal element an eigenvalue.

The initial step is that of proving the existence of at least one nonzero eigenvalue.

Theorem 6.4–A. *Suppose A compact, symmetric, and $A \neq 0$. Then either $\|A\|$ or $-\|A\|$ is an eigenvalue of A, and there is a corresponding eigenvector x such that $\|x\| = 1$ and $|(Ax, x)| = \|A\|$.*

PROOF. In view of (6.11–2) and (6.11–3) there exists a sequence $\{x_n\}$ such that $\|x_n\| = 1$ and $(Ax_n, x_n) \to \lambda$, where λ is real and $|\lambda| = \|A\|$. Now

$$\begin{aligned} 0 \leqslant \|Ax_n - \lambda x_n\|^2 &= \|Ax_n\|^2 - 2\lambda(Ax_n, x_n) + \lambda^2\|x_n\|^2 \\ &\leqslant \|A\|^2 - 2\lambda(Ax_n, x_n) + \lambda^2. \end{aligned}$$

But then we see that $Ax_n - \lambda x_n \to 0$. Since A is compact, $\{Ax_n\}$ contains a convergent subsequence, which we denote by $\{Ay_k\}$, $\{y_k\}$ being a subsequence of $\{x_n\}$. The sequence $\{y_k\}$ is then convergent also, because $\lambda \neq 0$. Suppose $y_k \to x$ as $k \to \infty$. Then $\|x\| = 1$ and $Ay_k \to Ax$, whence $Ax = \lambda x$. Evidently $|(Ax, x)| = |\lambda|\, \|x\|^2 = \|A\|$, so the proof is complete.

We now apply Theorem 6.4–A repeatedly. Denote the eigenvalue and eigenvector of Theorem 6.4–A by λ_1 and x_1 respectively. Let $X = X_1$, and let $X_2 = \{x : (x, x_1) = 0\}$. Then X_2 is a subspace invariant under A, for $x \in X_2$ implies $(Ax, x_1) = (x, Ax_1) = (x, \lambda_1 x_1) = \lambda_1(x, x_1) = 0$. The restriction of A to X_2 is compact and symmetric. If the restriction is not the zero operator, we can assert the existence of λ_2 and x_2 such that $x_2 \in X_2$, $\|x_2\| = 1$, $Ax_2 = \lambda x_2$, and $|\lambda_2|$ is the norm of the restriction of A to X_2. Evidently $|\lambda_2| \leqslant |\lambda_1|$. Continuing in this way we obtain the nonzero eigenvalues $\lambda_1, \cdots, \lambda_n$, with corresponding eigenvectors $x_1, \cdots, x_n$ of unit norm. We also obtain $X_1, \cdots, X_{n+1}$, with X_{k+1} the set of elements of X_k which are orthogonal to $x_1, \cdots, x_k$. At each step $x_k \in X_k$ and $|\lambda_k|$ is the norm of the restriction of A to X_k, so $|\lambda_1| \geqslant |\lambda_2| \geqslant \cdots \geqslant |\lambda_n|$. The process stops with λ_n, x_n and X_{n+1} if and only if the restriction of A to X_{n+1} is 0. In that case the range of A lies in the linear manifold generated by $x_1, \cdots, x_n$. For, if $x \in X$, let

$$y_n = x - \sum_{k=1}^{n} (x, x_k)x_k. \tag{6.4–1}$$

Then $(y_n, x_i) = 0$ if $i = 1, \cdots, n$, so that $y_n \in X_{n+1}$, and therefore $Ay_n = 0$, or

$$Ax = \sum_{k=1}^{n} (x, x_k)Ax_k = \sum_{k=1}^{n} \lambda_k(x, x_k)x_k. \tag{6.4–2}$$

This situation may occur even if X is infinite dimensional. It will certainly occur eventually if X is finite dimensional, because $x_1, \cdots, x_n$ are linearly independent.

The foregoing considerations lead us to the statement of the fundamental theorem:

Theorem 6.4–B. *Suppose A compact, symmetric, and $A \neq 0$. The procedure described in the foregoing discussion yields a possibly terminating sequence of nonzero eigenvalues $\lambda_1, \lambda_2, \cdots$ and a corresponding orthonormal set of eigenvectors $x_1, x_2, \cdots$. If the sequences do not terminate, then $|\lambda_n| \to 0$. The expansion*

$$Ax = \sum(Ax, x_k)x_k = \sum\lambda_k(x, x_k)x_k \tag{6.4–3}$$

is valid for each $x \in X$, the summation being extended over the entire sequence, whether finite or infinite. Each nonzero eigenvalue of A occurs in the sequence $\{\lambda_n\}$. The eigenmanifold corresponding to a particular λ_i is finite dimensional and its dimension is exactly the number of times this particular eigenvalue is repeated in the sequence $\{\lambda_n\}$.

PROOF. Since $|\lambda_k| \geqslant |\lambda_{k+1}|$, we either have $\lambda_n \to 0$ or $|\lambda_n| \geqslant \epsilon > 0$ for some ϵ and all n. Suppose the latter and that the sequence is infinite. Then $\{x_n/\lambda_n\}$ is a bounded sequence, and $A(x_n/\lambda_n) = x_n$, so that $\{x_n\}$ must contain a convergent subsequence. This is impossible, for the orthonormality yields $\|x_n - x_m\|^2 = 2$. Hence $\lambda_n \to 0$ when the sequence is infinite. If the sequence of λ_k's terminates with λ_n, (6.4–3) is equivalent to (6.4–2). In the nonterminating case we define y_n by (6.4–1) and obtain

$$\|y_n\|^2 = \|x\|^2 - \sum_{k=1}^{n} |(x, x_k)|^2 \leqslant \|x\|^2.$$

Since $y_n \in X_{n+1}$ and $|\lambda_{n+1}|$ is the norm of the restriction of A to X_{n+1}, we have

$$\|Ay_n\| \leqslant |\lambda_{n+1}| \, \|y_n\| \leqslant |\lambda_{n+1}| \, \|x\|.$$

Hence $Ay_n \to 0$. But

$$Ay_n = Ax - \sum_{k=1}^{n} (x, x_k)Ax_k,$$

and so we obtain (6.4–3) [note that $Ax_k = \lambda_k x_k$ and $(Ax, x_k) = (x, Ax_k) = \lambda_k(x, x_k)$].

If λ is a nonzero eigenvalue of A which is not in the sequence $\{\lambda_k\}$, there is a corresponding eigenvector x of unit norm, and it must be orthogonal to x_n for every n, by Theorem 6.11–A. Then $Ax = 0$, by (6.4–3). This contradicts $Ax = \lambda x \neq 0$. An eigenvalue cannot be repeated infinitely often in the sequence $\{\lambda_n\}$, because $\lambda_n \to 0$. Suppose that λ_k occurs p times. Then the corresponding eigenmanifold contains an orthonormal set of p eigenvectors, and is therefore at least p-dimensional. It cannot be of dimension greater than p, for this would entail the existence of an x such that $Ax = \lambda_k x$, $\|x\| = 1$, and $(x, x_n) = 0$ for every n. But such a thing is impossible, by an argument given at the beginning of this paragraph. The proof of Theorem 6.4–B is now complete.

The next theorem describes the inverse of $\lambda - A$.

Theorem 6.4–C. *Let A, $\{\lambda_n\}$, $\{x_n\}$ be as in Theorem 6.4–B. Then, if $\lambda \neq 0$ and if $\lambda \neq \lambda_k$ for each k, $\lambda - A$ has a continuous inverse defined on all of X and given by $x = (\lambda - A)^{-1}y$, where*

$$x = \frac{1}{\lambda}y + \frac{1}{\lambda}\sum \lambda_k \frac{(y, x_k)}{\lambda - \lambda_k} x_k. \tag{6.4–4}$$

PROOF. We can discover the foregoing formula as follows. Suppose x and y given, such that $\lambda x - Ax = y$. Then $Ax = \lambda x - y$, and so from (6.4–3) we have

$$\lambda x - y = \sum \lambda_k(x, x_k)x_k.$$

We form the inner product with x_i and obtain

$$(\lambda x, x_i) - (y, x_i) = \lambda_i(x, x_i).$$

Thus

$$(x, x_i) = \frac{(y, x_i)}{\lambda - \lambda_i},$$

and so

$$\lambda x = y + \sum \lambda_k \frac{(y, x_k)}{\lambda - \lambda_k} x_k,$$

which gives (6.4–4). This shows that the solution of $(\lambda - A)x = y$ is unique, if it exists. On the other hand, if the series in (6.4–4) is convergent, the element x defined by (6.4–4) certainly satisfies $(\lambda - A)x = y$, for then

$$\lambda x - Ax = y + \sum \lambda_k \frac{(y, x_k)}{\lambda - \lambda_k} x_k - \frac{1}{\lambda} Ay - \frac{1}{\lambda} \sum \lambda_k \frac{(y, x_k)}{\lambda - \lambda_k} Ax_k.$$

We put $Ax_k = \lambda_k x_k$ in the last sum and use (6.4–3) with y in place of x; the result is $\lambda x - Ax = y$.

We now show that the series in (6.4–4) *does* converge, no matter how y is chosen. For this purpose let

$$\alpha = \sup_k \left| \frac{\lambda_k}{\lambda - \lambda_k} \right|, \qquad \beta = \sup_k \frac{1}{|\lambda - \lambda_k|}.$$

Also let

$$u_n = \sum_{k=1}^{n} \lambda_k \frac{(y, x_k)}{\lambda - \lambda_k} x_k, \qquad v_n = \sum_{k=1}^{n} \frac{(y, x_k)}{\lambda - \lambda_k} x_k.$$

Now, if $m < n$,

$$\|u_n - u_m\|^2 = \sum_{k=m+1}^{n} \left| \frac{\lambda_k}{\lambda - \lambda_k} \right|^2 |(y, x_k)|^2 \leqslant \alpha^2 \sum_{k=m+1}^{n} |(y, x_k)|^2.$$

Therefore $\{u_n\}$ is a Cauchy sequence, because $\sum |(y, x_k)|^2$ is convergent (see § 3.2). If X were complete, this would be enough for our purpose. If X is not complete we continue the argument as follows:

$$\|v_n\|^2 = \sum_{k=1}^{n} \frac{|(y, x_k)|^2}{|\lambda - \lambda_k|^2} \leqslant \beta^2 \sum_{k=1}^{n} |(y, x_k)|^2 \leqslant \beta^2 \|y\|^2,$$

so that $\{v_n\}$ is bounded. Now $Av_n = u_n$. Hence, the compactness of A shows that $\{u_n\}$ contains a convergent subsequence. Being a Cauchy sequence, $\{u_n\}$ must then be convergent to the same limit as the subsequence. Hence the series in (6.4–4) converges. We see from (6.4–4) that

$$\|x\| \leqslant \frac{1}{|\lambda|} \|y\| + \frac{1}{|\lambda|} \alpha \|y\|.$$

Thus we see that $(\lambda - A)^{-1}$ is continuous and defined on all of X, with

$$\|(\lambda - A)^{-1}\| \leqslant \frac{1}{|\lambda|} \left[1 + \sup_k \left| \frac{\lambda_k}{\lambda - \lambda_k} \right| \right]. \tag{6.4–5}$$

We round out the foregoing discussion by considering the null manifold and range of A. The situation is clearest if X is complete.

Theorem 6.4–D. (*a*) *Let A, $\{\lambda_n\}$ and $\{x_n\}$ be as in Theorem* 6.4–B, *and let M be the closed linear manifold generated by the eigenvectors $x_1, x_2, \cdots$.*

Then $M^{\perp} = \mathcal{N}(A)$. *Hence the orthonormal set* $\{x_n\}$ *is complete if and only if* 0 *is not an eigenvalue of* A. *When* X *is complete we have* $X = M \oplus \mathcal{N}(A)$.
(*b*) *Suppose* X *is complete. Then the range of* A *is composed of those elements* y *in* M *which are such that the series*

$$\sum \frac{(y, x_k)}{\lambda_k} x_k \tag{6.4–6}$$

is convergent.

PROOF. (*a*) It follows from (6.4–3) that $M^{\perp} \subset \mathcal{N}(A)$. On the other hand, $x \in \mathcal{N}(A)$ implies $(x, x_k) = \lambda_k^{-1}(x, Ax_k) = \lambda_k^{-1}(Ax, x_k) = 0$, so that $x \in M^{\perp}$. Hence $M^{\perp} = \mathcal{N}(A)$. The orthonormal set $\{x_n\}$ is complete if and only if $M^{\perp} = (0)$, which then means that 0 is not an eigenvalue of A. If X is complete, we have $X = M \oplus \mathcal{N}(A)$, by Theorem 4.82–A. For the proof of (*b*) suppose $Ax = y$. Then from (6.4–3) we see that y is in M. From the orthonormality it follows that $(y, x_i) = \lambda_i(x, x_i)$. We can write $x = u + v$, $u \in M$, $v \in \mathcal{N}(A)$. Then $(x, x_k) = (u, x_k)$, since $v \perp M$, and so (see Theorem 3.2–H)

$$u = \sum (u, x_k)x_k = \sum (x, x_k)x_k = \sum \frac{(y, x_k)}{\lambda_k} x_k,$$

the series necessarily being convergent if in fact it is infinite. Conversely, suppose $y \in M$ and that the series (6.4–6) is convergent, with u as its sum. Then

$$Au = \sum(y, x_k)x_k = y,$$

so that $y \in \mathcal{R}(A)$. This completes the proof.

We now examine formula (6.4–3) and express it in a somewhat different way, for the purpose of showing the relation between the formula and the general spectral representation theorem (of § 6.5) for self-adjoint operators. For the purpose of the following discussion it is assumed that X is complete.

The series (6.4–3) remains convergent, with the same sum, no matter how the terms are rearranged in order (see the proof of Theorem 3.2–H). It is convenient to rearrange the terms, if necessary, in such a way that all the terms for which λ_k has one particular value are brought so that they occur consecutively in the series. We shall now assume that the notation has been arranged so that this is true. For each λ_k let P_k be the operator defined by

$$P_k x = \sum_{\lambda_i = \lambda_k} (x, x_i)x_i.$$

Then $P_j = P_k$ if $\lambda_j = \lambda_k$. It is easy to verify that $P_jP_k = 0$ if $\lambda_j \neq \lambda_k$, that $P_k^2 = P_k$, and that P_k is symmetric. The series (6.4–3) can now be written

$$Ax = \sum{}' \lambda_k P_k x, \tag{6.4–7}$$

where the prime mark on the summation sign indicates that the sum is extended over the distinct values of λ_k. In like manner, (6.4–4) can be put in the form

$$(\lambda - A)^{-1}y = \frac{1}{\lambda}y + \frac{1}{\lambda}\sum{}' \frac{\lambda_k}{\lambda - \lambda_k} P_k y. \tag{6.4–8}$$

This equation shows that λ_k is a first-order pole of $(\lambda - A)^{-1}y$, the residue being $P_k y$. It is easy to verify that P_k is the projection associated with the spectral set (λ_k) (as in § 5.8).

If we use the decomposition $X = M \oplus \mathscr{N}(A)$ (see Theorem 6.4–D), and write $x = u + v$, $u \in M$, $v \in \mathscr{N}(A)$, we have $P_k x = P_k u$ and

$$u = \sum{}' P_k x.$$

If we define P_0 by $P_0 x = v$, we see that

$$x = \sum{}' P_k x + P_0 x. \tag{6.4–9}$$

The operator P_0 is a symmetric projection, and $P_0 P_k = P_k P_0 = 0$ if $k \neq 0$.

Now let us define a one-parameter family of operators E_λ as follows (λ real, x arbitrary):

$$\begin{cases} E_\lambda x = \displaystyle\sum_{\lambda_k \leqslant \lambda}{}' P_k x & \text{if } \lambda < 0 \\ E_\lambda x = x - \displaystyle\sum_{\lambda_k > \lambda}{}' P_k x & \text{if } \lambda \geqslant 0. \end{cases} \tag{6.4–10}$$

It is understood that the meaning of a sum is 0 if there exists no points λ_k satisfying the indicated inequality. By separate consideration of the cases $\lambda \leqslant \mu < 0$, $\lambda < 0 \leqslant \mu$, $0 \leqslant \lambda \leqslant \mu$, it is easy to verify that $E_\mu E_\lambda = E_\lambda E_\mu = E_\lambda$ if $\lambda \leqslant \mu$. Hence in particular $E_\lambda^2 = E_\lambda$. Moreover, E_λ is symmetric. The operator E_λ is continuous from the right, in the sense that

$$\lim_{\lambda \to \mu^+} E_\lambda x = E_\mu x, \tag{6.4–11}$$

where $\lambda \to \mu^+$ means that we consider values $\lambda > \mu$. This is clear if we examine closely the definition of E_λ. The only possible point of accumulation of the points λ_k is at 0. As we move to the right from a given point μ, E_λ does not change in value except as λ reaches one of the points λ_k or the point 0. Thus (6.4–11) is evident if $\mu \neq 0$, and for $\mu = 0$ it is true as a result of the convergence of the series defining E_0.

Since $\sigma(A)$ lies on $[m(A), M(A)]$, it is easy to see that $E_\lambda = 0$ if $\lambda < m(A)$ and $E_\lambda = I$ if $\lambda \geqslant M(A)$.

We shall denote $\lim\limits_{\lambda \to \mu^-} E_\lambda x$ by $E_{\mu-0}x$. It is easy to see from (6.4–10) that

$$E_\lambda - E_{\lambda-0} = P_k \qquad \text{if } \lambda = \lambda_k$$

and that

$$E_0 - E_{0-0} = P_0.$$

Formula (6.4–7) can now be written in the form

$$Ax = \int_\alpha^\beta \lambda \, dE_\lambda x, \tag{6.4–12}$$

where $[\alpha, \beta]$ is any closed interval such that $\alpha < m(A)$ and $M(A) \leqslant \beta$. The integral here is defined by the usual procedure for a Riemann-Stieltjes integral, involving sums of the type

$$\sum \mu_i [E(\mu_i) - E(\mu_{i-1})]x,$$

with $\alpha = \mu_0 < \mu_1 < \cdots < \mu_n = \beta$. (Here we have written $E(\mu_i)$ instead of E with subscript μ_i, to simplify printing.) Formula (6.4–8) can also be expressed in integral form. First we modify (6.4–8) by use of (6.4–9) with y in place of x. In this way we find

$$(\lambda - A)^{-1}y = \sum{}' \frac{1}{\lambda - \lambda_k} P_k y + \frac{1}{\lambda - 0} P_0 y.$$

Then we have

$$(\lambda - A)^{-1}y = \int_\alpha^\beta \frac{1}{\lambda - \mu} \, dE_\mu y.$$

The results of this section have, as their most important application, the theory of integral equations with symmetric (or Hermitian) kernels, where the spaces and the kernels are such that the corresponding integral operators are compact. Compact integral operators were discussed in § 5.5. For symmetry of the operator, the condition on the kernel is $\overline{k(s, t)} = k(t, s)$. For a classical exposition of the theory of symmetric integral equations see Courant and Hilbert, **1**, pages 104–118. For an historical account, with many references, see Hellinger and Toeplitz, **1**, part III.

PROBLEMS

1. If $\lambda = \lambda_j$ for some j, show that the range of $\lambda - A$ consists of all vectors orthogonal to the eigenmanifold corresponding to λ_j. For such a vector y the general solution of $(\lambda - A)x = y$ is

$$x = \frac{1}{\lambda} y + \frac{1}{\lambda} \sum_{\lambda_k \neq \lambda} \lambda_k \frac{(y, x_k)}{\lambda - \lambda_k} + w,$$

where w is an arbitrary element of the eigenmanifold corresponding to λ_j

2. Suppose X is complete. Let $\{x_n\}$ be an orthonormal set, and let $\{\lambda_n\}$ be any sequence of real numbers such that $\lambda_n \to 0$. Let A be defined by $Ax = \sum_{k=1}^{\infty} \lambda_k(x, x_k)x_k$. Then A is self-adjoint and compact. For the compactness, start with any bounded infinite sequence of vectors. Then the sequence contains a weakly convergent subsequence (see Theorem 4.41–B), which we denote by $\{y_n\}$. Suppose $\|y_n\| \leqslant C$. Then

$$\|Ay_n - Ay_m\|^2 = \sum_{k=1}^{\infty} |\lambda_k|^2 |(y_n - y_m, x_k)|^2.$$

If $\epsilon > 0$, choose p so that $|\lambda_k| < \epsilon$ if $k > p$. Then

$$\sum_{k=p+1}^{\infty} |\lambda_k|^2 |(y_n - y_m, x_k)|^2 \leqslant \epsilon^2 \sum_{k=p+1}^{\infty} |(y_n - y_m, x_k)|^2 \leqslant \epsilon^2 \|y_n - y_m\|^2 \leqslant 4C^2\epsilon^2.$$

The weak convergence then enables us to show that $\{Ay_n\}$ is convergent, whence A is compact.

3. Suppose X is a complex inner-product space (not necessarily complete). Suppose S and T are compact members of $[X]$ such that $ST = TS$ and $(Sx, y) = (x, Ty)$ for all x and y. Then, there exists a finite or infinite orthonormal set $\{u_n\}$ and a corresponding sequence $\{\lambda_n\}$ such that $\lambda_n \neq 0$, $Su_n = \lambda_n u_n$, $Tu_n = \bar{\lambda}_n u_n$, $\|y\|^2 = \Sigma_n |(y, u_n)|^2$ for each $y \in \mathscr{R}(S)$, and $\lambda_n \to 0$ if the sequences are infinite. Method: Put $S + T = 2A$, $S - T = 2iB$, and consider A, B. They are symmetric and compact, $AB = BA$, and $S = A + iB$.

The foregoing proposition can be used to prove the completeness of the orthonormal set $\{v_\lambda\}$, where $v_\lambda(t) = e^{i\lambda t}$ and λ takes on all real values, in the space of continuous almost-periodic functions, with

$$(x, y) = \lim_{h\to\infty} \frac{1}{h} \int_0^h x(t)\overline{y(t)}\, dt.$$

This is an important example of a non-separable incomplete inner-product space. For this application of the abstract proposition $Sx = y$ is taken to mean

$$y(s) = \lim_{h\to\infty} \frac{1}{h} \int_0^h z(s - t)x(t)\, dt,$$

and $Tx = y$ means

$$y(s) = \lim_{h\to\infty} \frac{1}{h} \int_0^h \overline{z(s - t)}x(t)\, dt,$$

where z is a fixed continuous almost-periodic function. In this case the orthonormal set $\{u_n\}$ is a set $\{v_{\nu(n)}\}$, where $\nu(n)$ runs through all the values of λ for which $(z, v_\lambda) \neq 0$. The λ_n corresponding to u_n is $(z, v_{\nu(n)})$. For details see Rellich, **1**, pages 351–355.

6.41 Symmetric Operators with Compact Resolvent

As in § 6.4, we assume that X is an inner-product space, not necessarily complete. We also assume that X is not finite dimensional.

Theorem 6.41–A. *Suppose that T is a symmetric linear operator with domain and range in X, and suppose that T^{-1} exists, belongs to $[X]$, and is compact. Let $\{\lambda_n\}$, $\{x_n\}$ be the sequences of eigenvalues and eigenvectors associated with $A = T^{-1}$, as explained in connection with Theorem* 6.4–B, *and let $\mu_n = 1/\lambda_n$. The sequence $\{\mu_n\}$ is infinite, and $|\mu_n| \to \infty$. The orthonormal set $\{x_n\}$ is complete, and*

$$x = \sum_{k=1}^{\infty} (x, x_k)x_k \tag{6.41–1}$$

for each $x \in \mathscr{D}(T)$. A point μ is in $\sigma(T)$ if and only if it is one of the μ_k's. We have $Tx_n = \mu_n x_n$. If μ is not in $\sigma(T)$,

$$(\mu - T)^{-1}y = \sum_{k=1}^{\infty} \frac{(y, x_k)}{\mu - \mu_k} x_k \tag{6.41–2}$$

for each $y \in X$. This inverse operator is compact.

PROOF. The symmetry of T implies that of A; hence we can apply the results of § 6.4. We observe that $Ax_n = \lambda_n x_n$ is equivalent to $x_n = \lambda_n Tx_n$, or $Tx_n = \mu_n x_n$. The orthonormal set $\{x_n\}$ is complete, by Theorem 6.4–D*a*, because 0 is not an eigenvalue of A (since A^{-1} exists). The orthonormal set must therefore be infinite, since X is infinite dimensional. Hence $|\mu_n| \to \infty$, for we know that $\lambda_n \to 0$. The range of A is the domain of T; since each $x \in \mathscr{D}(T)$ is of the form $x = Ay$, (6.41–1) is a consequence of (6.4–3). Next we show that

$$(\mu - T)^{-1} = \mu^{-1}A\left(A - \frac{1}{\mu}\right)^{-1} \tag{6.41–3}$$

if μ is different from 0 and all of the μ_n's. First suppose that $x \in \mathscr{D}(T)$ and $(\mu - T)x = y$. Then $\mu Ax - x = Ay$, $(A - \mu^{-1})x = \mu^{-1}Ay$, and $x = \mu^{-1}(A - \mu^{-1})^{-1}Ay$. On the other hand, if $y \in X$ and $x = \mu^{-1}A(A - \mu^{-1})^{-1}y$, then $x \in \mathscr{D}(T)$ and it is easily verified that $(\mu - T)x = (A - \mu^{-1})(A - \mu^{-1})^{-1}y = y$. Since A permutes with $(A - \mu^{-1})^{-1}$, this proves (6.41–3). The compactness of $(\mu - T)^{-1}$ follows from this formula (see problem 4, § 5.5). To obtain the formula (6.41–2) we use (6.4–4). First we have

$$\left(\frac{1}{\mu} - A\right)^{-1} y = \mu y + \mu^2 \sum_{k=1}^{\infty} \frac{(y, x_k)}{\mu_k - \mu} x_k.$$

Then

$$\frac{1}{\mu}A\left(A - \frac{1}{\mu}\right)^{-1} y = -Ay + \mu \sum_{k=1}^{\infty} \frac{(y, x_k)}{\mu_k(\mu - \mu_k)} x_k.$$

If we express Ay by (6.4–3) and simplify, we obtain (6.41–2) from this result and (6.41–3). This finishes the proof.

We observe that the operator T of Theorem 6.41–A is not continuous, because $\|x_n\| = 1$ and $\|Tx_n\| = \|\mu_n x_n\| = |\mu_n| \to \infty$. However, T is closed, because $T = A^{-1}$ and A is closed.

Theorem 6.41–A can be applied to certain kinds of differential equation problems. For instance, consider the type of second-order differential operator discussed in Example 2, § 6.11, with boundary conditions such that the operator is symmetric. We shall now denote the operator by T instead of A. Then $Tx = u$ means that $x \in \mathscr{D}(T)$ (which includes specification of the boundary conditions) and

$$-\frac{d}{dt}[p(t)x'(t)] + q(t)x(t) = u(t).$$

Also, $(\mu - T)x = y$ means $x \in \mathscr{D}(T)$ and

$$\frac{d}{dt}[p(t)x'(t)] + [\mu - q(t)]x(t) = y(t). \tag{6.41–3}$$

In order to be able to apply Theorem 6.41–A to this operator T, we have to show that T^{-1} exists and is a compact operator defined on all of X. This can be shown to be true in certain cases by construction of a Green's function $k(s, t)$. The service provided by the Green's function is that of being the kernel for an integral operator which turns out to be T^{-1}. If one can show that the Green's function exists and that the corresponding integral operator is compact, then Theorem 6.41–A is applicable. The Green's function is Hermitian (i.e., $\overline{k(t, s)} = k(s, t)$) as a consequence of the symmetry of the operator T.

There is not space in this book for a detailed treatment of boundary-value problems for second-order ordinary differential equations. For an extensive discussion, using the full power of spectral-theory methods in Hilbert space, see Stone, **1**, pages 448–530. Another useful discussion of differential operators from the Hilbert space viewpoint is to be found in Achieser and Glasmann, **1**, pages 304–356. For an extensive discussion of eigenfunction expansions associated with the equation (6.41–3) [for the special case in which $p(t) = 1$] see Titchmarsh, **2**. Titchmarsh does not employ abstract operator-theory methods; however, his use of contour integrals and the calculus of residues is closely related to the fact that $(\lambda - T)^{-1}$ is an analytic function of λ on the resolvent set of T.

It is not always true that T has a compact resolvent. When it does not, there may be continuous spectrum in addition to or in place of point spectrum, and then the series expansions (6.41–1), (6.41–2) are replaced by integral representations.

PROBLEM

Suppose that H is a continuous symmetric operator such that H^{-1} exists and belongs to $[X]$. Let T and $A = T^{-1}$ be as in Theorem 6.41–A. Let $B = HAH$ (a compact and symmetric operator), and let $\{\lambda_n\}$ be the nonzero eigenvalues and $\{x_n\}$ the corresponding eigenvectors associated with B, just as in § 6.4. Show that $(\mu_n H^2 - T)(H^{-1}x_n) = 0$, where $\mu_n = 1/\lambda_n$. The orthonormal set $\{x_n\}$ is complete, and $x = \sum_1^\infty (x, Hx_n)H^{-1}x_n$ for each $x \in \mathscr{D}(T)$. If μ is different from every μ_n, $\mu H^2 - T$ has a bounded inverse defined on all of X and given by

$$(\mu H^2 - T)^{-1}y = \sum_1^\infty \frac{(y, H^{-1}x_n)}{\mu - \mu_n} H^{-1}x_n.$$

To show this, show first that

$$(\mu H^2 - T)^{-1} = \frac{1}{\mu} H^{-1}\left(B - \frac{1}{\mu}\right)^{-1} HA = \frac{1}{\mu} AH\left(B - \frac{1}{\mu}\right)^{-1} H^{-1}$$

if $\mu \neq 0$.

This problem has an application to the boundary-value problem corresponding to the equation (6.41–3) with $\mu\rho(t)$ in place of μ, where $\rho(t)$ is a positive real function such that, if H maps $x(t)$ into $\sqrt{\rho(t)}x(t)$, then H and H^{-1} are bounded operators on X.

6.5 The Spectral Theorem for Bounded Self-adjoint Operators

In this section we consider an arbitrary bounded self-adjoint operator A on a complex Hilbert space X. These assumptions about A and X apply to all the theorems of this section. Our aim is to obtain generalizations of the formula (6.4–12) and to use the results to establish a useful operational calculus. The methods are quite different from those used in § 6.4, for now we do not assume that A is compact, and the spectrum can be much more complicated.

Before stating the first theorem we call the reader's attention to the definitions of $m(A)$ and $M(A)$ in (6.11–2). Since A is continuous, $m(A)$ and $M(A)$ are finite.

Theorem 6.5–A. *There exists a family of orthogonal projections E_λ defined for each real λ, with the properties*

a. $E_\lambda E_\mu = E_\mu E_\lambda = E_\lambda$ *if* $\lambda \leqslant \mu$.

b. $\lim_{\mu \to \lambda+} E_\mu x = E_\lambda x$.

c. $E_\lambda = 0$ *if* $\lambda < m(A)$, $E_\lambda = I$ *if* $M(A) \leqslant \lambda$.

d. $E_\lambda A = AE_\lambda$.

For each x and y in X, $(E_\lambda x, y)$ is of bounded variation as a function of λ, and

$$(p(A)x, y) = \int_\alpha^\beta p(\lambda)\, d(E_\lambda x, y) \tag{6.5–1}$$

for each polynomial $p(\lambda)$ with real coefficients. The integral is an ordinary Stieltjes integral over any interval $[\alpha, \beta]$ such that $\alpha < m(A)$, $M(A) \leqslant \beta$.

PROOF. Let $p(\lambda)$ be any polynomial with real coefficients. Then p(A) is self-adjoint, and therefore the spectral radius of $p(A)$ is $\|p(A)\|$ (Theorem 6.2–C). Now $\sigma[p(A)]$ is the set of values of $p(\lambda)$ for $\lambda \in \sigma(A)$ (Theorem 5.2–D); consequently

$$\|p(A)\| = \sup_{\lambda\in\sigma(A)} |p(\lambda)|. \tag{6.5–2}$$

Now let C be the real Banach space of real-valued continuous functions f defined on the closed interval $m(A) \leqslant \lambda \leqslant M(A)$, with $\|f\|$ defined as the maximum of $|f(\lambda)|$ on the interval. If we consider the polynomial p as a member of C, we evidently have $\|p(A)\| \leqslant \|p\|$, by (6.5–2) and Theorem 6.2–B. We can consider the class P of polynomials p with real coefficients as a subspace of C. Since any member of C can be approximated uniformly as closely as we please by a member of P (the Weierstrass theorem), the subspace P is dense in C.

Now consider any two vectors x, y, and let

$$L(p) = (p(A)x, y).$$

It is readily evident that L is a complex-valued linear functional defined on P and that

$$|L(p)| \leqslant \|p\|\, \|x\|\, \|y\|. \tag{6.5–3}$$

Thus L is continuous on P and may be extended in a unique way by continuity to give a continuous linear functional defined on C. We denote the extension by L also. We now refer to § 4.32 for the representation of continuous linear functionals on the space C. We write $L(f) = L_1(f) + iL_2(f)$, where L_1 and L_2 have real values. On applying Theorem 4.32–B to L_1 and L_2 and then combining results, we see that there exists a complex-valued function of bounded variation $V(\lambda; x, y)$, depending on x and y as parameters, such that

$$(p(A)x, y) = \int_{m(A)}^{M(A)} p(\lambda)\, dV(\lambda : x, y) \tag{6.5–4}$$

for each p in P. In order to have $V(\lambda; x, y)$ uniquely determined by the functional L, we agree to have $V(\lambda; x, y)$ normalized in the manner explained in § 4.32, i.e., $V[m(A); x, y] = 0$ and $V(\lambda; x, y) = V(\lambda + 0; x, y)$ if $m(A) < \lambda < M(A)$.

In several parts of subsequent arguments we shall need the following uniqueness principle: If v is a function of bounded variation on $[a, b]$, normalized as in § 4.32, and if $\int_a^b t^n\, dv(t) = 0$ for $n = 0, 1, 2, \cdots$, it follows that $\int_a^b x(t)\, dv(t) = 0$ for every continuous function. This depends on the Weierstrass theorem about approximating continuous functions uniformly by polynomials. Then $v \sim 0$ and hence $v(t) \equiv 0$ (see § 4.32).

The next step is to show that for each λ, $V(\lambda; x, y)$ is a continuous symmetric bilinear form in x and y, in the sense of § 6.1. If α is any scalar and $n = 0, 1, 2, \cdots$, we have

$$\int \lambda^n\, dV(\lambda; \alpha x, y) = (A^n \alpha x, y) = \alpha(A^n x, y) = \alpha \int \lambda^n\, dV(\lambda; x, y)$$
$$= \int \lambda^n\, d[\alpha V(\lambda; x, y)].$$

We omit the limits on the integral signs, for convenience. The uniqueness principle mentioned above now shows that $V(\lambda; \alpha x, y) = \alpha V(\lambda; x, y)$. Similar arguments show that $V(\lambda;\ x_1 + x_2,\ y) = V(\lambda;\ x_1,\ y) + V(\lambda;\ x_2, y)$ and $V(\lambda; y, x) = \overline{V(\lambda; x, y)}$. For this last relation we use the fact that A^n is symmetric. The inequality (6.5–3) enables us to assert that the total variation of $V(\lambda; x, y)$ does not exceed $2\|x\|\, \|y\|$, for the real and imaginary parts of V have variations equal to $\|L_1\|$ and $\|L_2\|$ respectively. Hence $|V(\lambda;\ x,\ y)| \leqslant 2\|x\|\, \|y\|$; this shows that the bilinear form is continuous. By Theorem 6.1–B we now see that for each λ on $[m(A), M(A)]$ there exists a bounded linear operator $F(\lambda)$ such that $V(\lambda; x, y) = (F(\lambda)x, y)$. Since the bilinear form is symmetric, so is $F(\lambda)$. The fact that $V[m(A); x, y] = 0$ implies that $F[m(A)] = 0$. If we put $p(\lambda) \equiv 1$ in (6.5–4), we see that $V[M(A); x, y] = (x, y)$. Hence $F[M(A)] = I$. Next we prove that

$$F(\lambda)F(\mu) = F(\mu)F(\lambda) = F(\lambda) \qquad \text{if } \lambda \leqslant \mu. \tag{6.5–5}$$

In particular, $[F(\lambda)]^2 = F(\lambda)$, so that $F(\lambda)$ is a symmetric (and orthogonal) projection. The proof involves use of the following easily proved theorem about Stieltjes integrals: *Suppose f and g are continuous and v is of bounded variation on $[a, b]$. Let*

$$u(t) = \int_a^t g(s)\, dv(s).$$

Then

$$\int_a^b f(t)\, du(t) = \int_a^b f(t)g(t)\, dv(t).$$

We also need to know that, if v is normalized, so is u. Now let

$$U(\lambda; x, y) = \int_a^\lambda \mu^m \, dV(\mu; x, y),$$

where we have set $a = m(A)$, $b = M(A)$. Then

$$\int_a^b \lambda^n \, dU(\lambda; x, y) = \int_a^b \lambda^{m+n} \, dV(\lambda; x, y) = (A^{m+n}x, y)$$

$$= (A^n x, A^m y) = \int_b^b \lambda^n \, dV(\lambda; x, A^m y).$$

By the uniqueness principle we have

$$U(\lambda; x, y) = V(\lambda; x, A^m y).$$

Now

$$V(\lambda; x, A^m y) = (F(\lambda)x, A^m y) = (A^m F(\lambda)x, y) = \int_a^b \mu^m \, dV[\mu; F(\lambda)x, y].$$

Consequently

$$\int_a^\lambda \mu^m \, d(F(\mu)x, y) = \int_a^b \mu^m \, d(F(\mu)F(\lambda)x, y).$$

The integral on the left can be written as

$$\int_a^b \mu^m \, dW(\mu; x, y),$$

where

$$W(\mu; x, y) = \begin{cases} (F(\mu)x, y) & \text{if } a \leqslant \mu \leqslant \lambda \\ (F(\lambda)x, y) & \text{if } \lambda \leqslant \mu \leqslant b. \end{cases}$$

Thus, using the uniqueness principle again, we find that $(F(\mu)F(\lambda)x, y) = (F(\nu)x, y)$, where $\nu = \min(\lambda, \mu)$. This implies the result (6.5–5).

Lastly, as regards $F(\lambda)$, we assert that $F(\mu)x \to F(\lambda)x$ as $\mu \to \lambda^+$ if $a < \lambda < b$. We write $F(\lambda + 0)x = \lim\limits_{\mu \to \lambda^+} F(\mu)x$, by definition, so that our assertion is

$$F(\lambda + 0)x = F(\lambda)x, \qquad a < \lambda < b. \tag{6.5–6}$$

To prove this we observe by (6.5–5) and Theorem 6.3–C that $F(\mu) - F(\lambda)$ is a symmetric projection if $\lambda < \mu$. Therefore

$$\|F(\mu)x - F(\lambda)x\|^2 = (F(\mu)x - F(\lambda)x, x) = (F(\mu)x, x) - (F(\lambda)x, x).$$

This expression approaches 0 as $\mu \to \lambda^+$ if $a < \lambda < b$, because $V(\lambda; x, y)$ is normalized. Thus (6.5–6) is proved. Likewise, if $a < \lambda_1 < \lambda_2$,

$$\|F(\lambda_2)x - F(\lambda_1)x\|^2 = (F(\lambda_2)x, x) - (F(\lambda_1)x, x) \to 0$$

as λ_1 and $\lambda_2 \to a^+$, because $(F(\lambda)x, x) = V(\lambda; x, x) \to V(a + 0; x, x)$ as $\lambda \to a^+$. Therefore $\lim\limits_{\lambda \to a^+} F(\lambda)x$ exists and defines an operator $F(a + 0)$. We now know that

$$(p(A)x, y) = \int_a^b p(\lambda)\, d(F(\lambda)x, y). \tag{6.5–7}$$

We define $E_\lambda = 0$ if $\lambda < a$, $E_\lambda = I$ if $b \leqslant \lambda$, and $E_\lambda = F(\lambda + 0)$ if $a \leqslant \lambda < b$. This makes $E_\lambda = F(\lambda)$ if $a < \lambda \leqslant b$. It is easy to verify that conditions a, b, and c in Theorem 6.5–A are fulfilled. Since E_λ may differ from $F(\lambda)$ at $\lambda = a$, we cannot replace $F(\lambda)$ by E_λ in (6.5–7). But, if $\alpha < a$ and $b \leqslant \beta$, simple calculations show that

$$\int_b^\beta p(\lambda)\, d(E_\lambda x, y) = 0$$

and that

$$\int_\alpha^a p(\lambda)\, d(E_\lambda x, y) + \int_a^b p(\lambda)\, d\{(E_\lambda x, y) - (F(\lambda)x, y)\} = 0.$$

Thus we obtain (6.5–1).

The fact that $AE_\mu = E_\mu A$ is easily deduced from (6.5–1). We have

$$(E_\mu Ax, y) = (Ax, E_\mu y) = \int_\alpha^\beta \lambda d(E_\lambda x, E_\mu y),$$

$$(AE_\mu x, y) = \int_\alpha^\beta \lambda\, d(E_\lambda E_\mu x, y) = (E_\mu Ax, y),$$

because $(E_\lambda E_\mu x, y) = (E_\mu E_\lambda x, y) = (E_\lambda x, E_\mu y)$. Thus $E_\mu A = AE_\mu$. Theorem 6.5–A is now proved. ■

We observe that $(E_\lambda x, x)$ is a nondecreasing function of λ. This is equivalent to $E_\lambda \leqslant E_\mu$ if $\lambda < \mu$, which is equivalent to the known relation $E_\lambda E_\mu = E_\lambda$, by Theorem 6.3–C. Just as in a previous argument about $F(\lambda)$, we can also conclude that $\lim\limits_{\lambda \to \mu^-} E_\lambda x$ exists and defines an operator $E_{\mu-0}$, which is evidently a symmetric projection.

The next theorem shows that (6.5–1) is matched by a direct formula for $p(A)$, without intervention of x and y.

Theorem 6.5–B. *The formula*

$$p(A) = \int_\alpha^\beta p(\lambda)\, dE_\lambda \tag{6.5–8}$$

holds, the integral on the right being defined in the usual way as a limit of Riemann-Stieltjes sums, with convergence in the norm topology of the operator space $[X]$.

PROOF. We recall that $\alpha < m(A)$, $M(A) \leqslant \beta$. Suppose that $\alpha = \lambda_0 < \lambda_1 < \cdots < \lambda_n = \beta$ and that $\lambda_{k-1} \leqslant \mu_k \leqslant \lambda_k$, $k = 1, 2, \cdots, n$. Let

$$B = \sum_{k=1}^{n} p(\mu_k)[E(\lambda_k) - E(\lambda_{k-1})].$$

Let $\epsilon_k = \max |p(\lambda) - p(\mu)|$ as λ and μ vary over $[\lambda_{k-1}, \lambda_k]$, and let $\epsilon = \max(\epsilon_1, \cdots, \epsilon_n)$. Now

$$(p(A)x, x) = \sum_{k=1}^{n} \int_{\lambda_{k-1}}^{\lambda_k} p(\lambda)\, d(E_\lambda x, x),$$

$$(Bx, x) = \sum_{k=1}^{n} \int_{\lambda_{k-1}}^{\lambda_k} p(\mu_k)\, d(E_\lambda x, x).$$

Since $(E_\lambda x, x)$ is nondecreasing, we see that

$$(p(A)x, x) - (Bx, x) \leqslant \sum_{k=1}^{n} \int_{\lambda_{k-1}}^{\lambda_k} \epsilon\, d(E_\lambda x, x) = \epsilon(x, x),$$

and hence that $M[p(A) - B] \leqslant \epsilon$ [see (6.11–2)]. In a similar way we see that $-\epsilon \leqslant m[p(A) - B]$. But then $\|p(A) - B\| \leqslant \epsilon$, by (6.11–3). Now $\epsilon \to 0$ as we carry out the usual limiting process in connection with a Stieltjes integral, because of the continuity of p. We see in this way that (6.5–8) is true in the sense asserted. ∎

The family of projections E_λ having the properties specified in Theorem 6.5–A is unique. This follows as a result of the uniqueness principle for normalized functions of bounded variation which was mentioned in the course of the proof of Theorem 6.5–A. It can even be shown, though we do not give the details, that, if G_λ is a family of symmetric projections satisfying conditions a, b, and c of Theorem 6.5–A and if the equation

$$(Ax, x) = \int_{\alpha}^{\beta} \lambda\, d(G_\lambda x, x)$$

holds for every x, then $G_\lambda = E_\lambda$ for every λ. Because of this uniqueness, the operator A fully determines the family E_λ (and is, of course, fully determined by the family). This family of projections is called the *resolution of the identity* corresponding to A.

Next we consider the integral

$$\int_{\alpha}^{\beta} f(\lambda)\, dE_\lambda, \tag{6.5–9}$$

where f is an arbitrary complex-valued continuous function defined on the interval $m(A) \leqslant \lambda \leqslant M(A)$. We extend the definition of f by setting

$f(\lambda) = f[m(A)]$ if $\alpha \leqslant \lambda \leqslant m(A)$ and $f(\lambda) = f[M(A)]$ if $M(A) \leqslant \lambda \leqslant \beta$. In order to show that the Stieltjes integral (6.5–9) exists (as a limit of sums in the topology of $[X]$), we can proceed as follows: First suppose that f has real values. Referring back to the proof of Theorem 6.5–A, we can now see that

$$L(f; x, y) = \int_\alpha^\beta f(\lambda)\, d(E_\lambda x, y),$$

and that this is a continuous symmetric bilinear form in x and y. Hence there exists a uniquely determined self-adjoint operator, which we denote by $f(A)$, such that

$$(f(A)x, y) = \int_\alpha^\beta f(\lambda)\, d(E_\lambda x, y).$$

It is now easy, just as in the proof of Theorem 6.5–B, to show that

$$f(A) = \int_\alpha^\beta f(\lambda)\, dE_\lambda. \tag{6.5–10}$$

If f is complex valued, the integral in (6.5–10) exists, as we see by separating f into its real and imaginary parts. In this case also we denote the value of the integral by $f(A)$.

Theorem 6.5–C. *The correspondence between the continuous function f and the operator $f(A)$ indicated in formula* (6.5–10) *has the properties:*

a. $(f + g)(A) = f(A) + g(A)$.
b. $(\alpha f)(A) = \alpha f(A)$.
c. $(fg)(A) = f(A)g(A)$.
d. $f(A)B = Bf(A)$ *if* $B \in [X]$ *and* $BE_\lambda = E_\lambda B$ *for every* λ.
e. $f(A)$ *is normal, with* $f(A)^*$ *corresponding to the function with values* $\overline{f(\lambda)}$.
f. $f(A)$ *is self-adjoint if* f *is real-valued, and* $f(A) \geqslant 0$ *if* $f(\lambda) \geqslant 0$ *for every* λ.
g. $\|f(A)\| \leqslant \max |f(\lambda)|$ $\{\lambda$ *varying over* $[m(A), M(A)]\}$.
h. $\|f(A)x\|^2 = \int_\alpha^\beta |f(\lambda)|^2 d\|E_\lambda x\|^2$.

PROOF. Properties *a* and *b* are obvious. To prove (*c*) we use an approximating sum for $f(A)$:

$$\sum_{k=1}^{n} f(\lambda_k)\, [E(\lambda_k) - E(\lambda_{k-1})],$$

and the exactly corresponding sum for $g(A)$. On multiplying the sums, the fact that

$$[E(\lambda_k) - E(\lambda_{k-1})]\,[E(\lambda_j) - E(\lambda_{j-1})] = \begin{cases} E(\lambda_k) - E(\lambda_{k-1}) & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases}$$

yields us the sums

$$\sum f(\lambda_k)g(\lambda_k)\,[E(\lambda_k) - E(\lambda_{k-1})].$$

The limit of this sum is evidently $(fg)(A)$; but it is also $f(A)g(A)$. Thus (*c*) is proved. Property *d* may also be proved by using the approximating sums. We leave (*e*) and (*f*) to the reader. For (*g*) let $\|f\| = \max |f(\lambda)|$, and suppose $\alpha = \lambda_0 < \lambda_1 < \cdots < \lambda_n = \beta$. Then

$$x = [E(\lambda_n) - E(\lambda_0)]x = \sum_{k=1}^{n} [E(\lambda_k) - E(\lambda_{k-1})]x,$$

and by the orthogonality relations among the projections $E(\lambda_k) - E(\lambda_{k-1})$, we find that

$$\|x\|^2 = \sum_{k=1}^{n} \|[E(\lambda_k) - E(\lambda_{k-1})]x\|^2. \tag{6.5–11}$$

Also,

$$\begin{aligned} \|f(A)x\|^2 &= \lim \left(\sum_k f(\lambda_k)[E(\lambda_k) - E(\lambda_{k-1})]x, \sum_j f(\lambda_j)[E(\lambda_j) - E(\lambda_{j-1})]x\right) \\ &= \lim \sum_k |f(\lambda_k)|^2\, \|[E(\lambda_k) - E(\lambda_{k-1})]x\|^2 \\ &\leqslant \|f\|^2\, \|x\|^2, \end{aligned}$$

by (6.5–11). This proves (*g*) and at the same time proves (*h*), for

$$\|[E(\lambda_k) - E(\lambda_{k-1})]x\|^2 = ([E(\lambda_k) - E(\lambda_{k-1})]x, x) = \|E(\lambda_k)x\|^2 - \|E(\lambda_{k-1})x\|^2.$$

We can also obtain *h* from *c* and *e*. This completes the proof.

Theorem 6.5–C exhibits a homomorphism of the algebra of continuous functions on $[m(A), M(A)]$ into the algebra $[X]$. As in the case of the homomorphism referred to in Theorem 5.6–A, the present homomorphism furnishes what may be called an operational calculus.

Next we shall see that the behaviour of E_λ as a function of λ is closely correlated with $\sigma(A)$.

Theorem 6.5–D. *If λ_0 is real, then $\lambda_0 \in \rho(A)$ if and only if there is some $\epsilon > 0$ such that E_λ is constant when $\lambda_0 - \epsilon \leqslant \lambda \leqslant \lambda_0 + \epsilon$.*

PROOF. We prove the "if" part first. Let $f(\lambda) = \lambda_0 - \lambda$, and define $g(\lambda) = (\lambda_0 - \lambda)^{-1}$ except when $\lambda_0 - \epsilon \leqslant \lambda \leqslant \lambda_0 + \epsilon$; the values of g on this latter interval can be assigned arbitrarily, except for the requirement that g be continuous. Then $f(\lambda)g(\lambda) = 1$ except on the interval $[\lambda_0 - \epsilon, \lambda_0 + \epsilon]$. Since E_λ is constant on this interval, it follows that

$$f(A)g(A) = \int_\alpha^\beta f(\lambda)g(\lambda)\, dE_\lambda = \int_\alpha^\beta dE_\lambda = I.$$

Therefore $g(A)$ is the inverse of $\lambda_0 - A = f(A)$. This implies that $\lambda_0 \in \rho(A)$.

For the "only if' part we can assume $m(A) < \lambda_0 < M(A)$, because of Theorem 6.2–B and the known facts about E_λ for $\lambda < m(A)$ and $M(A) < \lambda$. Suppose, no matter how small ϵ is, that there are points λ_1 and λ_2 in the interval $[\lambda_0 - \epsilon, \lambda_0 + \epsilon]$ such that $\lambda_1 < \lambda_2$ and $E(\lambda_1) \neq E(\lambda_2)$. Since $E(\lambda_2)E(\lambda_1) = E(\lambda_1)$, the range M_1 of $E(\lambda_1)$ is properly contained in the range M_2 of $E(\lambda_2)$, and we can choose an element y in $M_2 - M_1$ and orthogonal to M_1. Then $E(\lambda_2)y = y$ and $E(\lambda_1)y = 0$. If $\lambda \leqslant \lambda_1$, then $E_\lambda y = E_\lambda E(\lambda_1)y = 0$; if $\lambda_2 \leqslant \lambda$, then $E_\lambda y = E_\lambda E(\lambda_2)y = E(\lambda_2)y = y$. Thus, since $E_\lambda y$ is constant if $\lambda \leqslant \lambda_1$ and also if $\lambda_2 \leqslant \lambda$, part h of Theorem 6.5–C shows that

$$\|(\lambda_0 - A)y\|^2 = \int_{\lambda_1}^{\lambda_2} (\lambda_0 - \lambda)^2\, d\|E_\lambda y\|^2 \leqslant \epsilon^2\|y\|^2.$$

The last inequality follows from the fact that $(E_\lambda y, y) = \|E_\lambda y\|^2$ is nondecreasing and never exceeds $\|y\|^2$. We now see that

$$\inf_{\|x\|=1} \|(\lambda_0 - A)x\| = 0,$$

which implies that $\lambda_0 \in \sigma(A)$. This completes the proof of Theorem 6.5–D.

We see readily, somewhat as in the first part of the proof of Theorem 6.5–D, that the resolvent of A is given by

$$(\lambda_0 - A)^{-1} = \int_\alpha^\beta \frac{1}{\lambda_0 - \lambda}\, dE_\lambda \tag{6.5–12}$$

if λ_0 is not on the interval $[m(A), M(A)]$. If λ_0 is a point of $\rho(A)$ on the aforementioned interval, (6.5–12) is still valid; the singularity of $(\lambda_0 - \lambda)^{-1}$ at λ_0 causes no trouble because of the fact that E_λ is constant on a neighborhood of λ_0.

Theorem 6.5–E. *The point spectrum $P\sigma(A)$ consists of those points μ for which $E_\mu \neq E_{\mu-0}$. The corresponding eigenmanifold is then the range of the projection $E_\mu - E_{\mu-0}$. The continuous spectrum $C\sigma(A)$ consists of those points μ for which $E_\mu = E_{\mu-0}$ but which are such that E_λ is not constant in any neighborhood of μ.*

PROOF. (For the definition of $E_{\mu-0}$ see the remarks following the proof of Theorem 6.5–A.) From the fact that $E_\lambda E_\mu = E_\lambda$ if $\lambda \leqslant \mu$ we easily deduce that $E_\lambda E_{\mu-0} = E_{\mu-0}$ if $\mu \leqslant \lambda$ and $E_\lambda E_{\mu-0} = E_\lambda$ if $\lambda < \mu$. Hence also $E_{\mu-0}$ and $E_\mu - E_{\mu-0}$ are projections (see Theorem 6.3–C). Suppose $E_\mu \neq E_{\mu-0}$. Let $y = (E_\mu - E_{\mu-0})x$, with $y \neq 0$. The foregoing relations show that $E_\lambda y = 0$ if $\lambda < \mu$ and $E_\lambda y = y$ if $\mu \leqslant \lambda$. Then, using Theorem 6.5–C*h*, we have

$$\|(\mu - A)y\|^2 = \int_\alpha^\beta (\mu - \lambda)^2\, d\|E_\lambda y\|^2 = \int_\alpha^\mu (\mu - \lambda)^2\, d\|E_\lambda y\|^2 = 0.$$

The vanishing of the last integral depends on the fact that $(\mu - \lambda)^2 = 0$ at $\lambda = \mu$. We see that μ is an eigenvalue of A.

Suppose, conversely, that $\mu \in P\sigma(A)$, $y \neq 0$, $Ay = \mu y$. Then

$$0 = \int_\alpha^\beta (\mu - \lambda)^2\, d\|E_\lambda y\|^2.$$

We may suppose $M(A) < \beta$, since the integral is independent of β as long as $M(A) \leqslant \beta$. Now $\alpha < m(A) \leqslant \mu < \beta$, so we may choose $\epsilon > 0$ so that $\alpha < \mu - \epsilon$ and $\mu + \epsilon < \beta$. We use the facts that $(\mu - \lambda)^2 \geqslant 0$ and $\|E_\lambda y\|^2 = (E_\lambda y, y)$ is nondecreasing to conclude that

$$\int_\alpha^{\mu-\epsilon} (\mu - \lambda)^2\, d\|E_\lambda y\|^2 = 0.$$

This integral, however, is not smaller than

$$\epsilon^2(\|E_{\mu-\epsilon} y\|^2 - \|E_\alpha y\|^2) = \epsilon^2\|E_{\mu-\epsilon} y\|^2.$$

Therefore $E_{\mu-\epsilon} y = 0$, whence $E_{\mu-0} y = 0$. We also conclude that

$$\int_{\mu+\epsilon}^\beta (\mu - \lambda)^2\, d\|E_\lambda y\|^2 = 0,$$

and from this that

$$0 = \epsilon^2(\|E_\beta y\|^2 - \|E_{\mu+\epsilon} y\|^2) = \epsilon^2(\|y\|^2 - \|E_{\mu+\epsilon} y\|^2),$$

whence $E_{\mu+\epsilon} y = y$ and finally that $E_\mu y = y$. Therefore $E_\mu \neq E_{\mu-0}$. Also, $y = (E_\mu - E_{\mu-0})y$, so that the assertion of the theorem is proved as far as $P\sigma(A)$ is concerned. The part about $C\sigma(A)$ follows with the aid of Theorem 6.5–D.

The actual determination of the resolution of the identity for a given operator A is not an easy matter, in general. In some comparatively simple cases it may be inferred or conjectured from (6.5–1) or 6.5–4). This is so with problems 5 and 6 at the end of this section. A methodical procedure is furnished by the following formula. Suppose $\alpha < m(A)$ and $\epsilon > 0$. Then

$$\tfrac{1}{2}[(E_{\mu-0}x, y) + (E_\mu x, y)] = \lim_{\epsilon \to 0} \frac{1}{2\pi i} \int_\Gamma ((\lambda - A)^{-1}x, y)\, d\lambda, \tag{6.5–13}$$

where Γ is the polygonal line joining $\mu + i\epsilon$, $\alpha + i\epsilon$, $\alpha - i\epsilon$, and $\mu - i\epsilon$ in that order. See Dunford, 3, page 58. The proof of (6.5–13) depends on the use of (6.5–12).

PROBLEMS

1. Let X be a complete inner-product space, with either real or complex scalars. Let $\{A_n\}$ be a sequence of self-adjoint operators such that $A_n \leqslant A_{n+1}$ and $(A_n x, x) \leqslant \alpha\|x\|^2$ for each n, where α is a real constant. Then there exists a self-adjoint operator A such that $A_n x \to Ax$ for each x. A similar proposition is valid if all the inequalities are reversed. For the proof, first show that the sequence $\{\|A_n\|\}$ is bounded. Then apply (6.11–7) to $H = A_n - A_m$, where $n > m$, treating H like $A - \lambda$ in the proof of Theorem 6.2–B. The result is

$$\|A_n x - A_m x\|^4 \leqslant [(A_n x, x) - (A_m x, x)]k\|x\|^2,$$

where k is a constant. The rest of the argument is left to the reader.

2. Let X be as in problem 1. Let $\{A_n\}$ be a sequence of self-adjoint operators such that $A_n \leqslant A_{n+1}$ for each n, and let A be an operator such that $(A_n x, x) \to (Ax, x)$ for each x. Then $A_n x \to Ax$. Instead of $A_n \leqslant A_{n+1}$ we may have $A_{n+1} \leqslant A_n$, and the conclusion is the same. For the proof the first step is to prove that $\lim_{n\to\infty} (A_n x, y)$ exists for each x and y. Next, show that $\{\|A_n\|\}$ is bounded. Then problem 1 is applicable to obtain $A_n x \to Bx$ for some B and each x. Finally, $A = B$.

3. Let X, A and E_λ be as in Theorem 6.5–A. Then for each λ there exists a sequence p_n of polynomials such that $p_n(A)x \to E_\lambda x$ for each x. This is readily evident for $\lambda < m(A)$ or $M(A) \leqslant \lambda$. Suppose $\alpha < m(A) \leqslant \mu < M(A) \leqslant \beta$, $0 < h < \beta - \mu$, and define $f(\lambda) = 1$, $1 - h^{-1}(\lambda - \mu)$, or 0 according as $\alpha \leqslant \lambda \leqslant \mu$, $\mu \leqslant \lambda \leqslant \mu + h$, or $\mu + h \leqslant \lambda \leqslant \beta$. Then using (6.5–10) we find

$$(f(A)x, x) = (E_\mu x, x) + \int_\mu^{\mu+h} f(\lambda)\, d(E_\lambda x, x).$$

It follows from this and problem 2 that $f(A)x \to E_\mu x$ as $h \to 0$. If $h = 1/n$, let the corresponding f be f_n. Choose a polynomial p_n such that $|p_n(\lambda) - f_n(\lambda)| < 1/n$ if $\alpha \leqslant \lambda \leqslant \beta$. With the aid of Theorem 6.5–Cg it can then be shown that $p_n(A)x \to E_\mu x$.

4. Show from problem 3 that, if $B \in [X]$ and $AB = BA$, then $BE_\lambda = E_\lambda B$.

5. With $X = l^2$, let $Ax = y$ be defined by $\eta_i = \alpha_i \xi_i$, where α_i is real and $\sup_i |\alpha_i| < \infty$. Then $\sigma(A)$ is the closure of the set of the α_i's, and the α_i's form $P\sigma(A)$. The resolution of the identity corresponding to A is defined by $(E_\lambda x, y) = \sum_{\alpha_i \leqslant \lambda} \xi_i \bar{\eta}_i$. This means that the matrix representing E_λ is a diagonal matrix with 1 in the ith diagonal position if $\alpha_i \leqslant \lambda$, and 0 there otherwise.

6. Consider $X = L^2(a, b)$, where (a, b) is a finite interval. Let $Ax = y$ be defined by $y(t) = tx(t)$. Then $\sigma(A)$ is the closed interval $[a, b]$, and $P\sigma(A)$ is empty. The resolution of the identity corresponding to A is defined by $E_\lambda x = u$, where (for $a \leqslant \lambda \leqslant b$), $u(t) = x(t)$ if $a \leqslant t \leqslant \lambda$, $u(t) = 0$ if $\lambda < t \leqslant b$.

7. Consider $X = L^2(-\infty, \infty)$, and suppose $a(t) \in \mathscr{L}^1(-\infty, \infty)$. Suppose also that $a(t)$ is real and that $a(-t) = {}^0\, a(t)$. Define $Ax = y$ by

$$y(s) = \int_{-\infty}^{\infty} a(s - t)x(t)\,dt.$$

Then A is self-adjoint and $\|A\| \leqslant \int_{-\infty}^{\infty} |a(t)|\,dt$. Let

$$b(s) = \int_{-\infty}^{\infty} e^{-ist} a(t)\,dt.$$

Then $\sigma(A)$ is the closure of the range of b, and $\lambda \in P\sigma(A)$ if and only if $b(s) = \lambda$ on a set of positive measure. The resolvent of A can be expressed in terms of Fourier transforms and then the resolution of the identity can be computed from (6.5–13). See Dunford, **3**, pages 60–64, and Pollard, **1**.

6.6 Unitary Operators

Let X be a complete complex inner-product space.

Definition. An operator $U \in [X]$ is called *unitary* if $UU^* = U^*U = I$.

Note that a unitary operator is normal and that $U^* = U^{-1}$. We have $(Ux, Uy) = (x, U^*Uy) = (x, y)$ for every x and y. It follows that U maps X isometrically onto all of X.

Theorem 6.6–A. *If U is unitary, $\sigma(U)$ lies on the circle $|\lambda| = 1$.*

PROOF. Since U and U^* are isometric, we see that $\|U\| = \|U^*\| = 1$. Hence $\lambda \in \rho(U) \cap \rho(U^*)$ if $|\lambda| > 1$. We know $0 \in \rho(U)$, because $U^{-1} = U^*$. Suppose $0 < |\lambda| < 1$. Then $\lambda^{-1} \in \rho(U^*)$. Now $\lambda - U = \lambda U(U^* - \lambda^{-1})$, and it follows that $(\lambda - U)^{-1} = \lambda^{-1}(U^* - \lambda^{-1})^{-1}U^*$, so that $\lambda \in \rho(U)$. This completes the proof.

Example 1. With $X = L^2(a, b)$ and α real let $Ux = y$ mean $y(t) = e^{i\alpha t}x(t)$. Then U is unitary. The set $\sigma(U)$ consists of the closure of the set of values of $e^{i\alpha t}$, $a < t < b$. The interval may be infinite.

There is a spectral theorem for unitary operators, corresponding closely to Theorem 6.5–A. In order to state the theorem we consider trigonometric polynomials, i.e., functions of the form

$$p(e^{it}) = \sum_{k=-n}^{n} c_k e^{ikt}, \tag{6.6–1}$$

with the complex coefficients c_k, and arbitrary $n (n = 0, 1, 2, \cdots)$. We define

$$p(U) = \sum_{k=-n}^{n} c_k U^k, \tag{6.6–2}$$

where $U^k = (U^{-1})^{-k}$ if $k < 0$. Then

$$p(U)^* = \sum_{k=-n}^{n} \bar{c}_k U^{-k}.$$

It is clear that $p(U)$ is normal.

Theorem 6.6–B. *Corresponding to the unitary operator U there is a family of symmetric projections E_t such that*

a. $E_sE_t = E_tE_s = E_s$ *if* $s \leqslant t$.
b. $E_{t+0} = E_t$.
c. $E_t = 0$ *if* $t \leqslant 0$, $E_t = I$ *if* $2\pi \leqslant t$.
d. $E_tU = UE_t$.

For each x and y, (E_tx, y) is of bounded variation in t, and (E_tx, x) is a nondecreasing function of t. For each trigonometric polynomial $p(e^{it})$

$$(p(U)x, y) = \int_0^{2\pi} p(e^{it})\, d(E_tx, y). \tag{6.6–3}$$

PROOF. Let $P[0, 2\pi]$ be that subspace of the complex Banach space $C[0, 2\pi]$ which is obtained by selecting out those functions f such that $f(0) = f(2\pi)$. We need to know how to represent the normed conjugate of $P[0, 2\pi]$ as a subspace of the space $BV[0, 2\pi]$. By Theorems 4.3–A and 4.32–B, each continuous linear functional on $P[0, 2\pi]$ is representable in the form

$$L(f) = \int_0^{2\pi} f(t)\, dv(t), \tag{6.6–4}$$

where the total variation of v is $\|L\|$. Let us define an equivalence relation in $BV[0, 2\pi]$, $w_1 \sim w_2$ meaning that

$$\int_0^{2\pi} f(t)\, dw_1(t) = \int_0^{2\pi} f(t)\, dw_2(t)$$

for each $f \in P[0, 2\pi]$. It turns out that $w \sim 0$ means $w(0) = w(2\pi)$ and $w(t + 0) = w(t - 0) = w(0 + 0) = w(2\pi - 0)$ if $0 < t < 2\pi$. We shall now say that w is normalized (relative to $P[0, 2\pi]$) if $w(0) = 0$ and $w(t + 0) = w(t)$ when $0 \leqslant t < 2\pi$. Then, much as in § 4.32, it can be shown that the normed conjugate of $P[0, 2\pi]$ is congruent to the subspace of normalized members of $P[0, 2\pi]$, under the correspondence $L \leftrightarrow v$ exhibited in (6.6–4).

If $p(e^{it})$ is a trigonometric polynomial, it is an element of $P[0, 2\pi]$, and as such has a norm $\|p\| = \max |p(e^{it})|$. We wish to show that $\|p(U)\| \leqslant \|p\|$. For this purpose we define a function F by $F(\lambda) = p(\lambda)$

if $0 < |\lambda| < 2$, $F(\lambda) = 0$ if $3 < |\lambda|$. We can then compute $F(U)$ by the operational calculus based on formula (5.6–10), for $F \in \mathfrak{A}_\infty(U)$, by Theorem 6.6–A. It is necessary to integrate over two circles, say $|\lambda| = {}^1/_2$ and $|\lambda| = {}^3/_2$. For $|\lambda| > 1$ we have

$$(\lambda - U)^{-1} = \sum_0^\infty \frac{U^n}{\lambda^{n+1}};$$

for $|\lambda| < 1$ we use $\lambda - U = (\lambda U^{-1} - 1)U$ and find

$$(\lambda - U)^{-1} = -U^{-1} \sum_0^\infty \lambda^n U^{-n}.$$

It is then easily verified that $F(U)$ coincides with $p(U)$ as defined in (6.6–2). It then follows from Theorem 5.71–A that $\sigma[p(U)]$ is the set of values assumed by $p(\lambda)$ for $\lambda \in \sigma(U)$. The spectral radius of $p(U)$ is $\|p(U)\|$, by Theorem 6.2–E. It then follows that $\|p(U)\| \leqslant \|p\|$.

The proof now proceeds much as in the case of a self-adjoint operator. Since

$$|(p(U)x, y)| \leqslant \|p\| \, \|x\| \, \|y\|,$$

$(p(U)x, y)$ is a continuous linear functional on the trigonometric polynomials, viewed as a subspace of $P[0, 2\pi]$. Since this subspace is dense in $P[0, 2\pi]$, $(p(U)x, y)$ determines uniquely a continuous linear functional $L(f; x, y)$ on $P[0, 2\pi]$. Let $V(t; x, y)$ be the normalized function of bounded variation corresponding to this functional. Then

$$(p(U)x, y) = \int_0^{2\pi} p(e^{it})\, dV(t; x, y). \tag{6.6–5}$$

The normalization ensures $V(0; x, y) = 0$, and $V(2\pi; x, y) = (x, y)$ follows by putting $p(\lambda) \equiv 1$. We leave to the reader the proof that $V(t; x, y)$ is a continuous bilinear form. Then there exists a family of symmetric operators E_t, defined when $0 \leqslant t \leqslant 2\pi$, such that $V(t; x, y) = (E_t x, y)$. Evidently $E_0 = 0$ and $E_{2\pi} = I$. We define $E_t = 0$ if $t < 0$ and $E_t = I$ if $2\pi < t$. The rest of the proof is similar to the argument in § 6.5, and we leave it to the reader.

Next comes the counterpart of Theorem 6.5–B.

Theorem 6.6–C. *The formula*

$$p(U) = \int_0^{2\pi} p(e^{it})\, dE_t \tag{6.6–6}$$

holds, the integral being defined as a limit in the norm topology of the space $[X]$.

PROOF. It will suffice to prove (6.6–6) for the special case $p(e^{it}) = e^{int}$; the general case will then follow by linearity. Form a subdivision $0 = t_0 < t_1 < \cdots < t_m = 2\pi$ and choose arbitrary points s_k such that $t_{k-1} \leqslant s_k \leqslant t_k$. Let

$$B = \sum_{k=1}^{m} e^{ins_k}[E(t_k) - E(t_{k-1})],$$

and write $A = U^n - B$. Now $(U^n x, U^n x) = (x, x)$. The operator $P_k = E(t_k) - E(t_{k-1})$ is a projection, and $P_j P_k = 0$ if $j \neq k$. From this we see that

$$BB^* = B^*B = \sum_{k=1}^{m} |e^{ins_k}|^2 P_k = I,$$

so that B is unitary. Therefore

$$\|Ax\|^2 = 2(x, x) - (Bx, U^n x) - (U^n x, Bx). \tag{6.6–7}$$

Next,

$$(E_s x, U^n x) = \int_0^s e^{-int}\, d(E_t x, x). \tag{6.6–8}$$

This is because

$$(E_s x, U^n x) = (U^{-n} E_s x, x) = \int_0^{2\pi} e^{-int}\, d(E_t E_s x, x),$$

and we can use (*a*) of Theorem 6.6–B. We can now write

$$\begin{aligned}
(Bx, U^n x) &= \sum_{k=1}^{m} e^{ins_k}(E(t_k)x - E(t_{k-1})x, U^n x) \\
&= \sum_{k=1}^{m} e^{ins_k} \int_{t_{k-1}}^{t_k} d(E_s x, U^n x) \\
&= \sum_{k=1}^{m} \int_{t_{k-1}}^{t_k} e^{in(s_k - t)}\, d(E_t x, x).
\end{aligned}$$

We used (6.6–8) at the last step. We now have

$$(U^n x, Bx) + (Bx, U^n x) = 2 \sum_{k=1}^{m} \int_{t_{k-1}}^{t_k} \cos n(s_k - t)\, d(E_t x, x).$$

Therefore from (6.6–7) we see that

$$\|Ax\|^2 = 2 \sum_{k=1}^{m} \int_{t_{k-1}}^{t_k} [1 - \cos n(s_k - t)]\, d(E_t x, x).$$

If $\epsilon > 0$, we can make all the intervals (t_{k-1}, t_k) so short that $1 - \cos n(s_k - t) < \epsilon/2$ if $t_{k-1} \leqslant t \leqslant t_k$. Then

$$\|Ax\|^2 < \epsilon \sum_{k=1}^{m} \int_{t_{k-1}}^{t_k} d(E_t x, x) = \epsilon \|x\|^2.$$

This completes the proof.

There are developments parallel to the later part of § 6.5. To each $f \in P[0, 2\pi]$ there corresponds a uniquely determined operator U_f such that

$$(U_f x, y) = \int_0^{2\pi} f(t)\, d(E_t x, y),$$

$$\|U_f x\|^2 = \int_0^{2\pi} |f(t)|^2\, d(E_t x, x),$$

and

$$U_f = \int_0^{2\pi} f(t)\, dE_t.$$

We write U_f rather than $f(U)$, because in a formal sense U_f results from putting U in place of e^{it}, not in place of t [e.g., $U_f = 1 - U^{-2}$ if $f(t) = 1 - e^{-2it}$]. We have $\|U_f\| \leqslant \|f\|$, and the mapping $f \to U_f$ has the properties corresponding to *(a)*–*(d)* of Theorem 6.5–C. In particular, if $|\lambda| \neq 1$,

$$(\lambda - U)^{-1} = \int_0^{2\pi} \frac{1}{\lambda - e^{it}}\, dE_t. \tag{6.6–9}$$

A point e^{is} is in $\sigma(U)$ if and only if s is not interior to an interval of constancy of E_t; if $0 < t \leqslant 2\pi$, $e^{it} \in P\sigma(U)$ if and only if $E_{t-0} \neq E_t$. We omit the details.

The family E_t is called the resolution of the identity for U.

Example 2. The Fourier-Plancherel transform defines a unitary operator F in $L^2(-\infty, \infty)$. The definition of F is $Fx = y$, where

$$y(t) =^0 \frac{1}{\sqrt{2\pi}} \frac{d}{dt} \int_{-\infty}^{\infty} \frac{e^{-ist} - 1}{-is}\, x(s)\, ds.$$

An alternative formula is

$$y(t) =^0 \underset{\alpha \to \infty}{\text{l.i.m.}} \frac{1}{\sqrt{2\pi}} \int_{-\alpha}^{\alpha} e^{-ist} x(s)\, ds,$$

where l.i.m. means "limit in mean," i.e., the limit in the metric of $L^2(-\infty, \infty)$. It turns out that $F^4 = I$, and with the aid of this it is rather easy to show that $\sigma(F)$ consists of the four eigenvalues ± 1, $\pm i$. The resolution of the identity for F is $E_t = 0$ if $t < \pi/2$, $E_t = P_1$ if

$\pi/2 \leqslant t < \pi$, $E_t = P_1 + P_2$ if $\pi \leqslant t < 3\pi/2$, $E_t = P_1 + P_2 + P_3$ if $3\pi/2 \leqslant t < 2\pi$, and $E_t = I$ if $2\pi \leqslant t$, where

$$P_1 = \tfrac{1}{4}(I - iF - F^2 + iF^3), \qquad P_2 = \tfrac{1}{4}(I - F + F^2 - F^3),$$

$$P_3 = \tfrac{1}{4}(I + iF - F^2 - iF^3).$$

See Riesz and Nagy, **1**, pages 291–292.

PROBLEMS

1. If $A \in [X]$ and $\|Ax\| = \|x\|$ for every x, then A is called isometric. Show that $A^*A = I$ in this case. An isometric operator is not necessarily unitary. Consider $X = l^2$ and $A(\xi_1, \xi_2, \cdots) = (0, \xi_1, \xi_2, \cdots)$.

2. If A is isometric but not unitary, then $\mathscr{R}(A)$ is a proper closed subspace of X. If X is finite dimensional, every isometric operator is unitary.

3. Discuss $\sigma(U)$ and $(\lambda - U)^{-1}$ if $U(\xi_1, \xi_2, \cdots) = (\xi_2, \xi_1, \xi_4, \xi_3, \cdots)$, with $X = l^2$.

6.7 Unbounded Self-adjoint Operators

In this section we shall discuss briefly the adjoints of unbounded operators and certain facts about unbounded symmetric operators, especially unbounded self-adjoint operators. The purpose of the section is to orient the student who desires to study these matters further elsewhere. There is not space in this book for an extensive treatment of these topics. The subject matter leads on naturally into applications to symmetric differential operators, and there are important applications to quantum mechanics.

We assume that X is a complex Hilbert space. Some parts of the considerations can be developed for real spaces and also for incomplete spaces.

Let T be a linear operator with range in X and with domain dense in X; this assumption is retained in all the following discussion. If y, z is a pair such that $(Tx, y) = (x, z)$ when $x \in \mathscr{D}(T)$, we say $y \in \mathscr{D}(T^*)$ and $T^*y = z$. This defines T^*. We now describe things which can be proved, without giving the proofs. The adjoint operator T^* is closed. If $\mathscr{D}(T^*)$ is dense in X, $(T^*)^* \equiv T^{**}$ is defined and is an extension of T. This occurs if and only if T has a closed linear extension, and then T^{**} is the minimal such extension. Hence $T = T^{**}$ if T is closed.

If T is symmetric, T^* is an extension of T, and hence $\mathscr{D}(T^*)$ is dense in X. The operator T^{**} turns out to be symmetric, but T^* is symmetric if and only if $T^* = T^{**}$. In general T^* is a proper extension of T^{**}.

If T is symmetric, $P\sigma(T)$ and $C\sigma(T)$ are confined to the real axis, but there may be nonreal points in $R\sigma(T)$. However, if λ is nonreal and in $\sigma(T)$, all points on the same side of the real axis as λ are in $R\sigma(T)$. The relation $T^* = T^{**}$ holds if and only if $\sigma(T)$ is confined to the real axis.

Definition. T is called *self-adjoint* if $T = T^*$.

If T is self-adjoint, it is closed, and $\sigma(T)$ is confined to the real axis. Conversely, if T is closed and symmetric and if $\sigma(T)$ is confined to the real axis, T is self-adjoint.

If T is closed and symmetric, but not self-adjoint, the operators $T \pm i$ have bounded inverses, and hence the ranges of these operators are closed. The dimensions of $\mathscr{R}(T - i)^\perp$ and $\mathscr{R}(T + i)^\perp$ are called the *deficiency indices* of T.

Various investigations of symmetric operators may be made with the aid of operators called *Cayley transforms*. If T is symmetric, the Cayley transform of T is the operator

$$V = (T - i)(T + i)^{-1}.$$

The domain of V is $\mathscr{R}(T + i)$, and V maps this domain isometrically onto $\mathscr{R}(T - i)$; V is unitary if and only if T is self-adjoint.

When T is self-adjoint, we can use the spectral representation of its Cayley transform to obtain a spectral representation of T and a generalization of the results of § 6.5. Applying Theorem 6.6–B to V, let us write

$$(Vx, y) = \int_0^{2\pi} e^{i\theta}\, d(F_\theta x, y),$$

where F_θ is the resolution of the identity for V. The relation between V and T then permits us to deduce (after some details which we omit) the relation

$$(6.7\text{–}1) \qquad (Tx, y) = -\int_0^{2\pi} \operatorname{ctn}\frac{\theta}{2}\, d(F_\theta x, y), \qquad x \in \mathscr{D}(T),\ y \in X.$$

This integral is improper at 0 and 2π. We then define $E_\lambda = F_\theta$, where $\lambda = -\operatorname{ctn}(\theta/2)$. Then (6.7–1) becomes

$$(Tx, y) = \int_{-\infty}^{\infty} \lambda\, d(E_\lambda x, y).$$

The family of symmetric projections E_λ has properties much as in Theorem 6.5–A, but we have $E_\lambda x \to 0$ as $\lambda \to -\infty$, $E_\lambda x \to x$ as $\lambda \to +\infty$, in place of property c in this earlier theorem. The domain of T consists exactly of those x for which

$$\int_{-\infty}^{\infty} \lambda^2\, d(E_\lambda x, x)$$

is convergent; the value of the integral is then $\|Tx\|^2$. The behavior of E_λ near a particular λ_0 indicates the classification of this point in $\rho(T)$ or $\sigma(T)$ just as in the bounded case. Finally, an operational calculus can be developed, generalizing Theorem 6.5–C.

Example 1. The operator A of Example 1, § 6.11 is self-adjoint. Its spectrum is made up of the points $\lambda_n = n$, $n = 0, \pm 1, \pm 2, \cdots$; λ_n is an eigenvalue corresponding to the eigenvector u_n, where $u_n(s) = (2\pi)^{-1/2}e^{ins}$. In this case the resolution of the identity corresponding to A is given by

$$[E(\lambda_n) - E(\lambda_n - 0)]x = (x, u_n)u_n.$$

For $x \in \mathscr{D}(A)$ the formula

$$Ax = \sum_{-\infty}^{\infty} (Ax, u_n)u_n = \sum_{-\infty}^{\infty} n(x, u_n)u_n$$

is just the same as the formula

$$Ax = \int_{-\infty}^{\infty} \lambda \, dE_\lambda x,$$

and $x \in \mathscr{D}(A)$ is equivalent to the convergence of

$$\int_{-\infty}^{\infty} \lambda^2 \, d(E_\lambda x, x) = \sum_{-\infty}^{\infty} n^2|(x, u_n)|^2.$$

The reader should refer to the discussion of Example 2, § 5.7; there, however, E_n is the operator now denoted by $E(\lambda_n) - E(\lambda_n - 0)$. See also Stone, **1**, pages 428–435.

Example 2. Let $X = L^2(-\infty, \infty)$. Let $\mathscr{D}(T)$ be the subspace of X determined by those functions $x(t)$ which are absolutely continuous on every finite interval, with $x(t)$ and $x'(t)$ belonging to $\mathscr{L}^2(-\infty, \infty)$. Let $Tx = y$ mean $y(t) = ix'(t)$. The operator T is closely related to the operator S for which $Sx = y$ means $y(t) = tx(t)$, $\mathscr{D}(S)$ being the subset of X for which $x(t)$ and $tx(t)$ are both in $\mathscr{L}^2(-\infty, \infty)$. The operators S and T are self-adjoint, and $T = FSF^{-1}$, where F is the Fourier-Plancherel operator (see Example 2, § 6.6). The entire real axis belongs to $\sigma(T)$, and it is all continuous spectrum. The same is true for S. The resolution of the identity for S is given by $E_\lambda x = y$, where $y(t) = x(t)$ if $t \leqslant \lambda$, $y(t) = 0$ if $\lambda < t$. For T the resolution of the identity is given by

$$(E_\lambda - E_\mu)x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{i\lambda(s-t)} - e^{i\mu(s-t)}}{i(s - t)} \, ds.$$

Another item of interest is the following, which relates to the operational calculus as it applies to T: Suppose $f \in \mathscr{L}^2(-\infty, \infty)$, and let g be the Fourier-Plancherel transform of f. Then the operator $f(T)$ corresponding to f is given by $f(T)x = y$, where

$$y(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(t - s)x(s)\, ds.$$

For details see Achieser and Glasmann, **1**, pages 118–120 and 227–232 and Stone, **1**, pages 441–446.

SUGGESTED READING

Achieser and Glasmann, **1**, Chapters 4, 5, 6, 7.
Cooke, **1**, Chapters 4, 5.
Dunford, **3**, especially pp. 41–64.
Friedman, **1**.
Halmos, **3**, Chapter 2.
Hellinger, **1**.
Hellinger and Toeplitz, **1**.
Ljusternik and Sobolew, **1**, Chapter 5.
Nagy, **1**.
von Neumann, **1**.
Riesz and Nagy, **1**, Chapters 6, 7, 8, 9.
Schmeidler, **1**, Parts II, III.
Stone, **1**, Chapters 5, 6, 8, 9, 10.
Wintner, **1**, Chapters 4, 5, 6.
Zaanen, **1**, Chapter 12.

7

INTEGRATION AND LINEAR FUNCTIONALS

7.0 Explanatory Remarks

In the earlier parts of this book we have from time to time referred to the function classes $\mathscr{L}^p$ and the associated spaces L^p. The functions in question were assumed to be defined and measurable in the Lebesgue sense on some measurable set in Euclidean space of one or more dimensions. The attendant integrals were Lebesgue integrals. A knowledge of classical Lebesgue theory would be sufficient to cope with these previous references to measure and integration.

In modern analysis some form of general theory of measure and integration is essential for proper understanding of numerous developments. Classical theory of functions of a real variable is being swallowed up by topology and abstract functional analysis. It is not one of the objectives of this book to develop the generalization of the Lebesgue theory in detail from its axiomatic foundations. Instead, we assist the reader not already acquainted with such generalizations by outlining one form which the development may take, leaving him to read the details elsewhere. The close relationship between the theory of integration and linear spaces is stressed, and the use of integrals in the representation of certain linear functionals is developed in detail.

Modern forms of the theory of measure and integration play a central role in the theory of topological groups and elsewhere in topological algebra, notably in certain parts of the theory of Banach algebras. Generalizations of the spectral analysis of self-adjoint and normal operators in Hilbert space are prominent in this respect.

The sections prior to § 7.3 are expository, with few proofs given. In § 7.3 the space L^p is defined, and it is proved to be complete (Theorem 7.3–A). Theorem 7.3–D is an important theorem about approximation of elements of L^p by integrable simple functions. Theorem 7.3–E is a sort of converse of Hölder's inequality which is very important in proving the fundamental representation theorem for continuous linear functionals on L^p (Theorems 7.4–A and 7.4–B).

The next portion of the chapter (§§ 7.5–7.7) is devoted to the study of measures in locally compact Hausdorff spaces and the representation of linear functionals on certain spaces of continuous functions. If T is the Hausdorff space and $C_\infty(T)$ is the normed linear space of all real-valued continuous functions on T, with compact support, the relation between regular Borel measures on T and positive linear functionals on $C_\infty(T)$ is sketched, with reference to other sources for proofs. After a discussion of signed Borel measures and vector lattices, we come ultimately to the fundamental theorem (7.7–G) which identifies the normed conjugate of $C_\infty(T)$ with the vector lattice of finite regular signed Borel measures on T (i.e., on the σ-ring of Borel sets in T).

Finally, in § 7.8 and § 7.9 we obtain representation theorems (7.8–A and 7.9–A) for continuous linear functionals on spaces, respectively, of bounded and essentially bounded functions. These representation theorems involve integrals with respect to finitely additive set functions.

7.1 The Space $L(\mu)$

In defining the integral of a real-valued function with respect to a measure μ we follow the exposition of Halmos, **2**, Chapters II, IV, V. The reader will need to have some familiarity with general measure theory and with the general form of the theory of integration with respect to a postulated measure. Our primary aim in this section is to view the definition of the class of integrable functions as a process of completing a normed linear space. In order to assist the reader who is not thoroughly acquainted with the theory of measure and integration from the point of view taken here, we indicate the principal definitions and steps in the development of the theory. For proofs the reader can refer to the book of Halmos or to other standard texts.

Definition. A nonempty class $\boldsymbol{R}$ of sets is called a *ring* if $E \cup F$ and $E - F$ belong to $\boldsymbol{R}$ whenever E and F do. The ring is called a σ-ring if it contains the union of every countable collection of its members.

A ring contains the empty set, and it contains the union and intersection of any finite number of its members.

Definition. A *measure* is a function μ defined on a ring $\boldsymbol{R}$, the values of μ being either real numbers or $+\infty$, subject to the conditions

a. $\mu(E) \geqslant 0$ if $E \in \boldsymbol{R}$.
b. $\mu(\emptyset) = 0$.
c. If $\{E_n\}$ is a sequence of pairwise disjoint members of $\boldsymbol{R}$ whose union is in $\boldsymbol{R}$, then $\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n)$ (with the usual conventions about $+\infty$ in relation to the series).

A measure also has the following properties, as a consequence of the definition:

d. If $E, F \in \boldsymbol{R}$ and $E \subset F$, then $\mu(E) \leqslant \mu(F)$.
e. In the situation of (*d*), if $\mu(E) < \infty$, then $\mu(F - E) = \mu(F) - \mu(E)$.

Definition. If T is any nonempty set, if $\boldsymbol{S}$ is a σ-ring of subsets of T such that T is the union of all members of $\boldsymbol{S}$, and if μ is a measure defined on $\boldsymbol{S}$, we call $(T, \boldsymbol{S}, \mu)$ a measure space. The members of $\boldsymbol{S}$ are called measurable sets.

The set T itself need not belong to $\boldsymbol{S}$, and the complement of a measurable set need not be measurable.

For the development of the theory of integration based on a measure space it is not necessary to know where the measure came from. In various particular instances of the abstract general theory it is usually the case that a measure is constructed from some set function that has some, but not all, of the properties of a measure.

Example 1. If T is Euclidean space of n-dimensions, $\boldsymbol{S}$ is the class of Lebesgue measurable sets in T, and μ is Lebesgue measure, then $(T, \boldsymbol{S}, \mu)$ is a measure space. In this case $T \in \boldsymbol{S}$ and $\mu(T) = +\infty$. The ring $\boldsymbol{S}$ and the measure μ are constructed by starting with the definition and construction of an *outer measure.*

Example 2. Let T be the set of positive integers, $\boldsymbol{S}$ the class of all subsets of T, with $\mu(E)$ equal to the number of elements in E (either an integer or $+\infty$).

A measure space determines a class of functions called *integrable*, and with each such function x there is associated a real number, called its integral, and denoted by

$$\int_T x\,d\mu \qquad \text{or} \qquad \int x\,d\mu.$$

We frequently omit the symbol T on the integral sign, if the context is such that there is no ambiguity. In order to define the class of integrable functions we begin with the concept of a *measurable function*. This concept involves only T and S, not the measure μ.

First we must define "Borel set of real numbers." We give the definition so that it can be used for any locally compact Hausdorff space.

Definition. If C is the class of all compact subsets of a locally compact Hausdorff space, the elements of the smallest σ-ring containing C are called the *Borel sets* of the space. This smallest σ-ring is the intersection of all σ-rings which contain C.

For Euclidean space an equivalent definition of the Borel sets is obtained if we use the class of all open sets instead of the class of all compact sets.

Now consider functions x defined on T, with values $x(t)$ in the set consisting of the real numbers and the symbols $+\infty$, $-\infty$. We define $N(x)$ as the set of $t \in T$ such that $x(t) \neq 0$.

Definition. The function x is called measurable (with respect to S) if $N(x) \cap x^{-1}(M) \in S$ for each Borel set M of real numbers and if also the sets $\{t : x(t) = +\infty\}$, $\{t : x(t) = -\infty\}$ belong to S.

If T belongs to S and if all values of x are real (*not* $\pm\infty$), it can be shown that x is measurable if and only if $x^{-1}(M) \in S$ for every Borel set M.

This definition of measurability is equivalent to the definition of measurability in the classical Lebesgue theory for the case in which T is Euclidean space and μ is Lebesgue measure.

For the (T, S, μ) of Example 2, all functions are measurable.

If $E \in S$, the characteristic function of E is measurable. We denote it by x_E. For convenience in printing it is best to avoid subscripts on subscripts whenever possible; hence, we shall write $x_{E(n)}$ for the characteristic function of E_n.

Definition. The function x is called *simple* if there exists a finite collection of pairwise disjoint sets $E_1, \cdots, E_n$ in S such that the value of x on E_k is a real constant α_k (not $\pm\infty$) and the value of x is 0 on $T - (E_1 \cup \cdots \cup E_n)$. Then $x = \alpha_1 x_{E(1)} + \cdots + \alpha_n x_{E(n)}$.

The simple functions form a real linear space. The absolute value of a simple function is simple, and the product of two simple functions is simple. It is frequently useful to know that, if x is measurable, there exists a sequence $\{x_n\}$ of simple functions such that $x_n(t) \to x(t)$ for each $t \in T$. If $x(t) \geqslant 0$ for each t, it may be arranged to have $x_n(t) \geqslant 0$ and $x_n(t) \leqslant x_{n+1}(t)$.

Definition. A simple function x is called *integrable* if $\mu[N(x)] < \infty$. If x is an integrable simple function with the distinct nonzero values $\alpha_1, \cdots, \alpha_n$, assumed on sets $E_1, \cdots, E_n$ respectively, the integral of x is defined to be

$$\int x\, d\mu = \alpha_1\mu(E_1) + \cdots + \alpha_n\mu(E_n). \tag{7.1–1}$$

If $x(t) \equiv 0$, we define the integral of x to be 0.

It is easy to see that the integrable simple functions form a linear space and that the integral is a linear functional on this space. It is also easy to see that

$$\left|\int x\, d\mu\right| \leqslant \int |x|\, d\mu. \tag{7.1–2}$$

If x is an integrable simple function and $E \in S$, the product function $x\chi_E$ is also simple and integrable; we define the integral of x over the set E as

$$\int_E x\, d\mu = \int_T x\chi_E\, d\mu. \tag{7.1–3}$$

If x and y are measurable functions, we write $x(t) =^0 y(t)$ and say that $x(t) = y(t)$ almost everywhere (abbreviated as a.e.) provided that the measure of the set $\{t : x(t) \neq y(t)\}$ is 0. It is clear that $=^0$ is an equivalence relation on the class of integrable simple functions and that we can form a linear space whose elements are equivalence classes of integrable simple functions. Equivalent functions have equal integrals. A function x is equivalent to the identically zero function if and only if $\int |x|\, d\mu = 0$. It then follows easily that the space of equivalence classes becomes a normed linear space if we define the norm of an element of the space to be $\int |x|\, d\mu$, where x is any member of the equivalence-class element. All of this depends on just a few of the elementary properties of integrals of integrable simple functions.

Let us denote by W the normed linear space of equivalence classes of integrable simple functions, as described in the foregoing paragraph. For convenience we ignore the notational distinction between an integrable simple function and the equivalence class to which it belongs. It is clear from (7.1–2) that the integral is a continuous linear functional on W. If $\hat{W}$ is the completion of W, the integral, as a continuous linear functional on W, has a uniquely determined extension which is a continuous linear functional on $\hat{W}$. Our aim is to show how $\hat{W}$ can be regarded as a space whose elements are equivalence classes of functions more general than the integrable simple functions. These functions are what we call *integrable* (but not necessarily simple) functions, and the extended linear functional defines the integral for such functions.

The essential concept which is needed at this stage is that of *convergence in measure.*

Definition. A sequence $\{x_n\}$ of measurable functions whose values are a.e. finite (i.e., real, not $\pm\infty$) is called a Cauchy sequence in measure if for each $\epsilon > 0$ the measure of the set $\{t:|x_n(t) - x_m(t)| \geqslant \epsilon\}$ converges to 0 as m and n become infinite. If x is a function of the same type as the x_n's, the sequence is said to converge to x in measure if for each $\epsilon > 0$ the measure of $\{t:|x_n(t) - x(t)| \geqslant \epsilon\}$ converges to 0 as $n \to \infty$. For brevity we write

$$x_n - x_m \xrightarrow{\mu} 0 \qquad \text{and} \qquad x_n \xrightarrow{\mu} x$$

respectively in these situations.

It is not a complicated matter to show that, if $x_n \xrightarrow{\mu} x$ and $x_n \xrightarrow{\mu} y$, then $x(t) =^0 y(t)$. A sequence is Cauchy in measure if and only if it has a limit in measure.

Suppose now that $\{x_n\}$ is a sequence of integrable simple functions such that

$$\int |x_n - x_m|\, d\mu \to 0 \qquad \text{as } m \text{ and } n \to \infty.$$

If $E_{mn} = \{t:|x_n(t) - x_m(t)| \geqslant \epsilon\}$, it follows that

$$\int |x_n - x_m|\, d\mu \geqslant \int_{E_{mn}} |x_n - x_m|\, d\mu \geqslant \epsilon\mu(E_{mn}),$$

and so $\{x_n\}$ is Cauchy in measure. Hence there exists a measurable and a.e. finite valued x such that $x_n \xrightarrow{\mu} x$. Next, suppose that $\{y_n\}$ is a sequence of integrable simple functions such that

$$\int |y_n - y_m|\, d\mu \to 0 \qquad \text{as } m, n \to \infty,$$

and hence, for a certain y, $y_n \xrightarrow{\mu} y$. Suppose also that $\int |x_n - y_n|\, d\mu \to 0$. Then it is easy to show that $x(t) =^0 y(t)$. Conversely (and this is more difficult to prove), if $\{x_n\}$ and $\{y_n\}$ are sequences of integrable simple functions such that $\int |x_n - x_m|\, d\mu \to 0$, $\int |y_n - y_m|\, d\mu \to 0$, $x_n \xrightarrow{\mu} x$, $y_n \xrightarrow{\mu} y$, and $x(t) =^0 y(t)$, then

$$\int |x_n - y_n|\, d\mu \to 0. \tag{7.1–4}$$

We shall prove this converse assertion.

First we note that to each $\epsilon > 0$ corresponds a $\delta > 0$ such that for all n

$$\int_F |x_n|\,d\mu < \epsilon \qquad \text{if } \mu(F) < \delta.$$

A similar thing holds for $\{y_n\}$. See Halmos, **2**, Theorem C, page 100. Now, for any given $\epsilon > 0$, let

$$E_n = \{t:|x_n(t) - y_n(t)| \geqslant \epsilon\}, \qquad F_n = \{t:|x_n(t) - x(t)| \geqslant \epsilon/3\},$$
$$G_n = \{t:|y_n(t) - y(t)| \geqslant \epsilon/3\}, \qquad H = \{t:|x(t) - y(t)| \geqslant \epsilon/3\}.$$

From the inequality

$$|x_n(t) - y_n(t)| \leqslant |x_n(t) - x(t)| + |x(t) - y(t)| + |y(t) - y_n(t)|$$

it is clear that $E_n \subset F_n \cup G_n \cup H$, and hence that $\mu(E_n) \leqslant \mu(F_n) + \mu(G_n) + \mu(H)$. But $\mu(H) = 0$ and $\mu(F_n) \to 0$, $\mu(G_n) \to 0$, whence $\mu(E_n) \to 0$. If now E is any measurable set of finite measure and $D_n = E \cap E_n$, we have

$$\int_E |x_n - y_n|\,d\mu \leqslant \int_{E-E_n} |x_n - y_n|\,d\mu + \int_{D_n} |x_n|\,d\mu + \int_{D_n} |y_n|\,d\mu.$$

The first term on the right does not exceed $\epsilon\mu(E)$. Then, since $\mu(D_n) \leqslant \mu(E_n) \to 0$, we see that

$$\int_E |x_n - y_n|\,d\mu \to 0$$

for every measurable set E of finite measure.

Next, if E is *any* measurable set, the limit

$$\nu(E) = \lim_{n\to\infty} \int_E |x_n - y_n|\,d\mu$$

exists, as is evident from the inequality

$$\left|\int_E (|x_n - y_n| - |x_m - y_m|)\,d\mu\right| \leqslant \int |x_n - x_m|\,d\mu + \int |y_n - y_m|\,d\mu.$$

It is easily proved that ν is a countably additive set function on S; we omit the details of this part of the argument. Since $\nu(E) = 0$ if $\mu(E)$ is finite, it follows that $\nu(E) = 0$ if E is any countable union of sets of finite measure. In particular, $\nu(E) = 0$ if E is the union of all the sets $N(x_n)$, $N(y_n)$, $n = 1, 2, \cdots$. With this last meaning of E we have

$$\int_T |x_n - y_n|\,d\mu = \int_E |x_n - y_n|\,d\mu \to \nu(E) = 0.$$

Thus (7.1–4) is proved.

A consequence of the conclusion just reached is that, with $\{x_n\}$ and $\{y_n\}$ as given,

$$\lim_{n\to\infty} \int x_n \, d\mu = \lim_{n\to\infty} \int y_n \, d\mu. \tag{7.1–5}$$

The limits exist, for

$$\left| \int x_n \, d\mu - \int x_m \, d\mu \right| \leqslant \int |x_n - x_m| \, d\mu \to 0,$$

and likewise for the integrals involving $\{y_n\}$. The truth of (7.1–5) follows from

$$\left| \int x_n \, d\mu - \int y_n \, d\mu \right| \leqslant \int |x_n - y_n| \, d\mu \to 0.$$

This brings us to the general definition of an integral.

Definition. An a.e. finite-valued measurable function x is called integrable if there exists a sequence of integrable simple functions $\{x_n\}$ such that

$$\int |x_n - x_m| \, d\mu \to 0 \quad \text{and} \quad x_n \xrightarrow{\mu} x. \tag{7.1–6}$$

The integral of x is then defined as

$$\int x \, d\mu = \lim_{n\to\infty} \int x_n \, d\mu. \tag{7.1–7}$$

The foregoing considerations justify this definition by showing that the limit in (7.1–7) exists and has the same value for any choice of $\{x_n\}$ satisfying (7.1–6). Moreover, equivalent integrable functions have equal integrals. It is easily verified that the integral in (7.1–7) has the same value as that defined earlier if x happens to be simple and integrable; in that case we can take $x_n = x$ for every n.

Suppose that x is an integrable function which originates from the sequence of integrable simple functions $\{x_n\}$ as in (7.1–6). Then $|x|$ originates in the same manner from $\{|x_n|\}$; therefore $|x|$ is integrable, and

$$\int |x_n| \, d\mu \to \int |x| \, d\mu. \tag{7.1–8}$$

We can conclude from this that $\int |x| \, d\mu = 0$ implies $x(t) =^0 0$. For, it certainly implies $x_n \xrightarrow{\mu} 0$; since $x_n \xrightarrow{\mu} x$ also, we have $x(t) =^0 0$.

We denote the class of integrable functions by $\mathscr{L}(\mu)$. The set of classes of equivalent integrable functions is denoted by $L(\mu)$; $L(\mu)$ is a normed linear space with $\int |x| \, d\mu$ as the norm of the element corresponding to the integrable function x.

We leave it for the reader to verify in detail the correlation between the process of completing the space W and our procedure in defining $\mathscr{L}(\mu)$ and $L(\mu)$. This correlation shows that $L(\mu)$ can be identified with the completion $\hat{W}$. We know that W is dense in $\hat{W}$ and that $\hat{W}$ is complete. In terms of functions, these facts can be stated as follows: *If x is integrable, there exists a sequence $\{x_n\}$ of integrable simple functions such that $\int|x_n - x|\, d\mu \to 0$. If $\{y_n\}$ is a sequence of integrable functions such that $\int|y_n - y_m|\, d\mu \to 0$, there exists an integrable function y such that $\int|y_n - y|\, d\mu \to 0$.*

By using Lebesgue's theorem on dominated convergence (Halmos, **2**, Theorem D, page 110), it is easy to prove the following: *If x is measurable and $|x(t)| \leqslant y(t)$ a.e., where y is integrable, then x is integrable. In particular, if constant functions are integrable, then every bounded measurable function is integrable.* A constant function on T is integrable if and only if T is measurable and $\mu(T)$ is finite.

PROBLEMS

1. With (T, S, μ) as in Example 2, show that convergence in measure is the same as uniform convergence.

2. With (T, S, μ) as in Example 2 and functions x on T expressed as sequences $x = \{\xi_n\}$, discuss the meaning of "simple function," "integrable simple function," and "integrable function." Show that x is integrable if and only if $\sum_1^\infty |\xi_n| < \infty$. What is $\int x\, d\mu$?

3. If (T, S, μ) is a measure space and z is a complex-valued function defined on T, let $z = x + iy$, where x and y are real valued. We say that z is measurable if x and y are both measurable, and likewise for integrability. If z is integrable, we define its integral as

$$\int z\, d\mu = \int x\, d\mu + i \int y\, d\mu.$$

Show that $|z|$ is integrable if z is. Show that, if z is integrable, there exists a sequence $\{z_n\}$ of integrable simple functions such that $\int|z_n - z|\, d\mu \to 0$. Discuss $\mathscr{L}(\mu)$ and $L(\mu)$ for complex functions, and show that $L(\mu)$ is a complex Banach space.

7.2 Signed and Complex Measures

It is useful to generalize the concept of a measure by allowing the function μ to assume values which are negative or complex. However, to avoid the ambiguity of $\infty - \infty$, it is necessary to require that μ can take on at most one of the values $+\infty$, $-\infty$.

Definition. Suppose T is a nonempty set, $\boldsymbol{S}$ is a σ-ring of subsets of T such that T is the union of all the members of $\boldsymbol{S}$, and μ is a function on $\boldsymbol{S}$ with values in a set consisting of all real numbers and *just one* of the symbols $+\infty$, $-\infty$. Then, if $\mu(\emptyset) = 0$ and if μ is countably additive, μ is called a *signed measure*. Being countably additive means that, for each sequence $\{E_n\}$ of pairwise disjoint members of $\boldsymbol{S}$ such that $\bigcup_{n=1}^{\infty} E_n \in \boldsymbol{S}$, then $\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n)$.

When we speak of a measure without a qualifying adjective such as "signed" or "complex" we mean a measure in the sense defined at the beginning of § 7.1.

With a signed measure μ we can associate two measures μ^+, μ^- defined as follows:

$$\begin{cases} \mu^+(E) = \sup \mu(F) \\ \mu^-(E) = -\inf \mu(F), \end{cases} \tag{7.2–1}$$

where F varies over all members of $\boldsymbol{S}$ such that $F \subset E$. It turns out that

$$\mu(E) = \mu^+(E) - \mu^-(E). \tag{7.2–2}$$

This formula expresses what is called the *Jordan decomposition* of μ (see Halmos, **2**, pages 120–123, especially problem 3, page 123; see also Saks, **1**, pages 10–11). If μ happens to be a measure (i.e., if $\mu(E) \geqslant 0$ always), then $\mu = \mu^+$.

We define a function $|\mu|$ by the formula

$$|\mu|(E) = \mu^+(E) + \mu^-(E). \tag{7.2–3}$$

This function is a measure on $\boldsymbol{S}$. It is called the *total variation* measure associated with μ. Observe that $|\mu|(E)$ is in general not the same as $|\mu(E)|$. It can be proved that $|\mu|$ is obtainable from μ in a different way. If $E \in \boldsymbol{S}$, let π_E denote any finite collection of pairwise disjoint sets $E_1, \cdots, E_n$ from $\boldsymbol{S}$ such that $E_k \subset E$. Then it can be shown that

$$|\mu|(E) = \sup_{\pi_E} \sum_{k=1}^{n} |\mu(E_k)|. \tag{7.2–4}$$

A signed measure is called *finite* if all the values of μ are real numbers (not $+\infty$ or $-\infty$). When μ is finite, $|\mu|$ is finite also, and the values of $|\mu|$ are bounded, as a result of the fact that $\boldsymbol{S}$ is a σ-ring. The class of finite signed measures becomes a real linear space in an obvious way. The function defined by

$$\|\mu\| = \sup_{E \in \boldsymbol{S}} |\mu|(E) \tag{7.2–5}$$

is a norm on this space of measures, and it is easy to prove that the space of measures is complete (see problem 1).

Let μ be a signed measure, and let x be a real-valued function which is defined on T and measurable with respect to the σ-ring $\boldsymbol{S}$. If x is integrable with respect to $|\mu|$, it is readily seen that x is also integrable with respect to μ^+ and μ^-. We then define the integral of x with respect to μ by the formula

$$\int x\,d\mu = \int x\,d\mu^+ - \int x\,d\mu^-. \tag{7.2–6}$$

It is easily shown that

$$\left|\int x\,d\mu\right| \leqslant \int |x|\,d|\mu|. \tag{7.2–7}$$

Definition. If T and $\boldsymbol{S}$ are as in the definition of a signed measure, by a *complex measure* we understand a function $\mu = \mu_1 + i\mu_2$, where μ_1 and μ_2 are signed measures on $\boldsymbol{S}$.

For a complex measure μ we can *define* the associated total variation $|\mu|$ by (7.2–4). It turns out that $|\mu|$ is a measure. If μ_1 and μ_2 are finite we say that μ is finite. Finite complex measures form a complex Banach space with the norm defined by (7.2–5).

We can define integrals with respect to a complex measure μ. Let z be a complex-valued function defined on T, and let $z = x + iy$, x and y having real values. Suppose that x and y are integrable with respect to $|\mu|$. Then they are also integrable with respect to $|\mu_1|$ and $|\mu_2|$, and we define

$$\int z\,d\mu = \int x\,d\mu_1 - \int y\,d\mu_2 + i\int y\,d\mu_1 + i\int x\,d\mu_2. \tag{7.2–8}$$

The inequality (7.2–7) is valid with μ a complex measure and x replaced by a complex-valued function z. Likewise, the result of problem 2 extends to the case of complex functions and complex measures.

PROBLEMS

1. Let $\{\mu_n\}$ be a sequence of finite signed measures such that $\|\mu_n - \mu_m\| \to 0$. Show that $\lim\limits_{n\to\infty} \mu_n(E) = \mu(E)$ exists uniformly with respect to E, that μ is a finite signed measure, and that $\|\mu_n - \mu\| \to 0$.

2. If μ is a signed measure and $|\mu|(E) < \infty$, show that

$$|\mu|(E) = \sup_{|x|\leqslant 1} \left|\int_E x\,d\mu\right|$$

(supremum over all real-valued $|\mu|$-integrable functions x for which $|x(t)| \leqslant 1$ on T). Use (7.2–7) and (7.2–4). The integral over E is defined by (7.1–3).

7.21 The Radon-Nikodym Theorem

Suppose that (T, S, μ) is a measure space and that x is an integrable function with values which are either real numbers or $\pm\infty$ (x is necessarily a.e. finite valued). Let

$$\nu(E) = \int_E x\, d\mu, \qquad E \in S. \tag{7.21-1}$$

It is known from the theory of integration that ν is a finite signed measure. It is also known that $\mu(E) = 0$ implies $\nu(E) = 0$.

We now consider a converse situation.

Definition. A signed measure ν on the σ-ring S is said to be *absolutely continuous* with respect to the measure μ on S if $\nu(E) = 0$ whenever $\mu(E) = 0$. If ν is a finite signed measure, it can be shown that this requirement is equivalent to the requirement that to each $\epsilon > 0$ correspond a $\delta > 0$ such that $|\nu(E)| < \epsilon$ if $\mu(E) < \delta$.

In § 7.4 we shall need the following theorem:

Theorem 7.21-A. (Radon-Nikodym). *Suppose that T itself is measurable and that it is the union of a countable family of measurable sets of finite measure. Let ν be a finite signed measure which is absolutely continuous with respect to the measure μ. Then, there exists an integrable function x such that $\nu(E)$ is given by* (7.21-1).

For a proof of this theorem the reader is referred to Halmos, **2**, Theorem B and proof, pages 128–130. We have stated the theorem in the form appropriate for our use, which is not the most general known form of the theorem.

7.3 The Real Space $L^p(\mu)$

Let (T, S, μ) be a measure space as in § 7.1. If $p > 0$, a measurable function x is said to belong to $\mathscr{L}^p(\mu)$ if $|x|^p$ is integrable. We can argue just as in Example 7, § 1.2, to show that $\mathscr{L}^p(\mu)$ contains the sum of any two of its members. We denote by $L^p(\mu)$ the linear space whose elements are the classes of equivalent members of $\mathscr{L}^p(\mu)$. If $p \geqslant 1$, the space $L^p(\mu)$ becomes a normed linear space with

$$\|x\|_p = \left(\int |x|^p\, d\mu\right)^{1/p} \tag{7.3-1}$$

as the norm of the element represented by the function x. The inequalities of Hölder and Minkowski are valid for these general $L^p(\mu)$ spaces (see Halmos, **2**, pages 175–176). We drop the superscript altogether if $p = 1$. From now on we confine attention to values of p such that $p \geqslant 1$. For convenience we write $\mathscr{L}^p$ and L^p instead of $\mathscr{L}^p(\mu)$ and $L^p(\mu)$.

Theorem 7.3–A. *If* $p \geqslant 1$, *the space* L^p *is a Banach space* (i.e., *it is complete*).

PROOF. For $p = 1$ the result is known from § 7.1. For the general case we have to prove the following: Suppose $\{x_n\}$ is a sequence of members of $\mathscr{L}^p$ such that

$$\int |x_n - x_m|^p \, d\mu \to 0 \qquad \text{as } m \text{ and } n \to \infty. \tag{7.3–2}$$

Then there exists a function $x \in \mathscr{L}^p$ such that

$$\int |x_n - x|^p \, d\mu \to 0 \qquad \text{as } n \to \infty. \tag{7.3–3}$$

Now, (7.3–2) implies that $\{x_n\}$ is a Cauchy sequence in measure. For, if $\epsilon > 0$, let $E_{mn} = \{t: |x_n(t) - x_m(t)| \geqslant \epsilon\}$. Then

$$\int |x_n - x_m|^p \, d\mu \geqslant \epsilon^p \mu(E_{mn}),$$

and the result follows. Hence there is a measurable function x such that $\{x_n\}$ converges in measure to x. Also, a subsequence converges a.e. to $x(t)$. Denote this subsequence of $\{x_n\}$ by $\{u_n\}$. If $\epsilon > 0$, choose $K = K(\epsilon)$ so that

$$\int |u_j - u_k|^p \, d\mu < \epsilon \qquad \text{if } j \text{ and } k \geqslant K.$$

For a fixed $j \geqslant K$ let $y_k = |u_j - u_k|^p$, $y = |u_j - x|^p$. Then $y_k(t) \to y(t)$ a.e., and

$$\int y_k \, d\mu < \epsilon$$

if $k \geqslant K$. Hence, by Fatou's lemma (Halmos, **2**, Theorem F, page 113), y is integrable and

$$\int y \, d\mu \leqslant \epsilon.$$

This shows that $u_j - x \in \mathscr{L}^p$ if $j \geqslant K$. Thus $x \in \mathscr{L}^p$ also. We also conclude that

$$\int |u_j - x|^p \, d\mu \to 0 \qquad \textit{as } j \to \infty.$$

From this, using (7.3–2) and Minkowski's inequality, we deduce (7.3–3), by the same argument as in showing that, if a subsequence of a Cauchy sequence has a certain limit, the whole sequence has this limit.

In working with the space L^p it is often convenient to know that the elements of L^p which correspond to integrable simple functions are everywhere dense in the space. We prove this by an argument in several stages. (For $p = 1$, of course, this result is implicit in the very definition of $\mathscr{L}$). Observe that a simple function is in $\mathscr{L}^p$ if and only if it is integrable.

In what follows the reader will need to recall that $N(x) = \{t : x(t) \neq 0\}$.

Lemma 7.3–B. *If $x \in \mathscr{L}^p$, there exists a sequence $\{x_n\}$ with $x_n \in \mathscr{L}^p$, $\mu[N(x_n)] < \infty$, $|x_n(t)| \leqslant |x(t)|$ for every n and t, and $\|x_n - x\| \to 0$.*

PROOF. It is known (Halmos, **2**, Theorem F, page 105) that for any integrable x the set $N(x)$ is contained in the union of a countable family of measurable sets of finite measure. Since $N(x)$ is measurable, we can then express $N(x)$ as the union of a countable family of pairwise disjoint sets of finite measure. Now $N(x) = N(|x|^p)$, and so this applies to x if $x \in \mathscr{L}^p$. Hence let $N(x) = \bigcup_{n=1}^{\infty} E_n$ be such a representation of $N(x)$. Now define x_n by

$$x_n(t) = \begin{cases} x(t) & \text{on } E_1 \cup \cdots \cup E_n \\ 0 & \text{elsewhere in } T. \end{cases}$$

Let $N(|x_n - x|) = F_n$; note that $F_n \subset N(x) - (E_1 \cup \cdots \cup E_n)$. Now $x_n \in \mathscr{L}^p$, because $|x_n| \leqslant |x|$. Then

$$\int |x_n - x|^p \, d\mu = \int_{F_n} |x|^p \, d\mu \leqslant \sum_{k=n+1}^{\infty} \int_{E_k} |x|^p \, d\mu.$$

But

$$\int |x|^p \, d\mu = \int_{N(x)} |x|^p \, d\mu = \sum_{k=1}^{\infty} \int_{E_k} |x|^p \, d\mu,$$

the series being convergent, since an integral is a countably additive set function. Therefore it is clear that $\|x_n - x\|_p \to 0$. Also, $N(x_n) \subset E_1 \cup \cdots \cup E_n$, so that $\mu[N(x_n)] < \infty$. This finishes the proof.

Lemma 7.3–C. *If $x \in \mathscr{L}^p$ and $\epsilon > 0$, there exists $y \in \mathscr{L}^p$ such that $\mu[N(y)] < \infty$, $|y(t)| \leqslant |x(t)|$ for each t, the values $|y(t)|$ are bounded, and $\|y - x\|_p < \epsilon$.*

PROOF. We first choose $z \in \mathscr{L}^p$ as one of the x_n's in Lemma 7.3–B so that $\|x - z\|_p < \epsilon/2$. We shall now show how to choose $y \in \mathscr{L}^p$ so that $\mu[N(y)] < \infty$, $|y| \leqslant |z|$, the values of $|y|$ are bounded, and $\|z - y\|_p < \epsilon/2$.

The y will then have the properties we seek, by Minkowski's inequality. Let us define z_n by

$$z_n(t) = \begin{cases} z(t) & \text{if } |z(t)| \leqslant n \\ n \operatorname{sgn} z(t) & \text{if } |z(t)| > n. \end{cases}$$

(See § 4.32 for the meaning of sgn.) Then z_n is measurable, $|z_n| \leqslant n$, and $|z_n| \leqslant |z|$. Note that $N(z_n) = N(z)$, whence $\mu[N(z_n)] < \infty$. Let E_n be the set where $|z(t)| > n$. Then

$$\int |z|^p \, d\mu \geqslant \int_{E_n} |z|^p \, d\mu \geqslant n^p \mu(E_n),$$

and so $\mu(E_n) \to 0$. Then

$$\int |z_n - z|^p \, d\mu = \int_{E_n} |z_n - z|^p \, d\mu \leqslant \int_{E_n} (|z_n| + |z|)^p \, d\mu \leqslant 2^p \int_{E_n} |z|^p \, d\mu.$$

From this it follows that $\|z_n - z\|_p \to 0$. We can choose y as z_n for some n large enough to make $\|z_n - z\|_p < \epsilon/2$, and the proof is complete.

Theorem 7.3–D. *If $x \in \mathscr{L}^p$ and $\epsilon > 0$, there exists an integrable simple function y such that $|y(t)| \leqslant |x(t)| + \epsilon$ for all values of t, and $\|x - y\|_p < \epsilon$.*

PROOF. By Lemma 7.3–C we can choose $z \in \mathscr{L}^p$ so that $\mu[N(z)] < \infty$, z is bounded, $|z| \leqslant |x|$, and $\|x - z\|_p < \epsilon/2$. Then we show how to choose an integrable simple function y such that $|y| \leqslant |z| + \epsilon$ and $\|z - y\|_p < \epsilon/2$. This will be enough to prove the theorem.

Let $A = \sup |z(t)|$. Divide the interval $(-A, A)$ of the real axis into n equal parts by points $\alpha_0, \alpha_1, \cdots, \alpha_n$, where $-A = \alpha_0 < \alpha_1 < \cdots < \alpha_n = A$. We define z_n as follows: $z_n(t) = 0$ if $z(t) = 0$, and, if $z(t) \neq 0$, then

$$z_n(t) = \begin{cases} \alpha_0 & \text{if } \alpha_0 \leqslant z(t) \leqslant \alpha_1 \\ \alpha_k & \text{if } \alpha_k < z(t) \leqslant \alpha_{k+1}, \quad k = 1, \cdots, n-1. \end{cases}$$

Then $N(z_n) \subset N(z)$, and z is an integrable simple function. It is evident that $|z_n(t) - z(t)| \leqslant 2A/n$ for every t. Then

$$\int |z_n - z|^p \, d\mu = \int_{N(z)} |z_n - z|^p \, d\mu \leqslant (2A/n)^p \, \mu[N(z)]$$

and $|z_n(t)| \leqslant |z(t)| + (2A/n)$. By taking n large enough the choice $y = z_n$ will be satisfactory.

We shall presently need the following theorem, which is a sort of converse of Hölder's inequality; we adopt the notation of § 4.32 as regards p and p'.

Theorem 7.3–E. *Suppose* $1 < p < \infty$. *Suppose that* T *itself is measurable and that it is expressible as a countable union of measurable sets of finite measure. Let* y *be a real-valued function which is defined a.e. on* T. *Suppose the product function* xy *is integrable whenever* x *is a bounded and measurable function on* T *such that* $\mu[N(x)] < \infty$. *Moreover, suppose that there exists a real constant* M *such that*

$$\left|\int xy\, d\mu\right| \leqslant M\|x\|_p \tag{7.3–4}$$

for every such x. *Then* $y \in \mathscr{L}^{p'}$ *and*

$$\int |y|^{p'}\, d\mu \leqslant M^{p'}. \tag{7.3–5}$$

PROOF. Let E be any measurable set of finite measure, and let x_E be its characteristic function. Then $x_E y$ is integrable, by hypothesis. Let $T = \bigcup\limits_{n=1}^{\infty} E_n$, where the sets E_n are measurable, pairwise disjoint, and of finite measure. Let $F(n) = E_1 \cup \cdots \cup E_n$, and define $y_n(t) = y(t)$ on $F(n)$, $y(t) = 0$ on $T - F(n)$. Then $y_n = yx_{F(n)}$ is measurable, and $y_n(t) \to y(t)$ a.e., so that y is measurable. In what follows we shall prove that $x_E|y|^{p'}$ is integrable for every E of finite measure, and that

$$\int x_E|y|^{p'}\, d\mu \leqslant M^{p'}. \tag{7.3–6}$$

It will then follow by Fatou's lemma that $y \in \mathscr{L}^{p'}$ and that (7.3–5) holds, for, in the notation used above,

$$x_{F(n)}(t)|y(t)|^{p'} \to |y(t)|^{p'} \quad \text{a.e.}$$

If E is any set of finite measure, define $x_n(t) = 0$ if $t \in T - E$, and, if $t \in E$, define

$$x_n(t) = \begin{cases} |y(t)|^{p'-1} \operatorname{sgn} y(t) & \text{if } |y(t)|^{p'-1} \leqslant n^{1/p} \\ n^{1/p} \operatorname{sgn} y(t) & \text{if } |y(t)|^{p'-1} > n^{1/p}. \end{cases}$$

Then x_n is measurable, $|x_n| \leqslant n^{1/p}$, $N(x_n) = N(y) \cap E$, and so $\mu[N(x_n)] < \infty$. Also, if $t \in E$,

$$|x_n(t)|^p = \begin{cases} |y(t)|^{p'} & \text{if } |y(t)|^{p'} \leqslant n \\ n & \text{if } |y(t)|^{p'} > n, \end{cases}$$

and

$$x_n(t)y(t) = \begin{cases} |y(t)|^{p'} & \text{if } |y(t)|^{p'} \leqslant n \\ n^{1/p}|y(t)| & \text{if } |y(t)|^{p'} > n. \end{cases}$$

Note that $x_n(t)y(t) \geqslant |x_n(t)|^p$ for every t. Therefore, by (7.3–4),

$$\int |x_n|^p \, d\mu \leqslant \int x_n y \, d\mu \leqslant M\Big(\int |x_n|^p \, d\mu\Big)^{1/p},$$

and it follows that

$$\Big(\int |x_n|^p \, d\mu\Big)^{1/p'} \leqslant M.$$

But $|x_n(t)|^p \to \chi_E(t)|y(t)|^{p'}$ a.e. Therefore, by Fatou's lemma, $\chi_E|y|^{p'}$ is integrable and (7.3–6) holds. In view of remarks made earlier, this completes the proof.

PROBLEMS

1. In Example 6, § 3.11, we defined the function class $\mathscr{L}^\infty$ and the space L^∞ for the case of Lebesgue measure on the real axis. These definitions can be extended at once, and we obtain $\mathscr{L}^\infty(\mu)$ and $L^\infty(\mu)$ in relation to an arbitrary measure space (T, S, μ). Show that $L^\infty(\mu)$ is complete.

2. Let the assumptions on y be as in Theorem 7.3–E, except that now we assume $p = 1$. Then we can conclude that $y \in \mathscr{L}^\infty(\mu)$ and that $\sup^0 |y(t)| \leqslant M$.

3. State and prove a theorem corresponding to Theorem 7.3–E for the case $p = \infty$, $p' = 1$, the conclusion being that y is integrable and $\int |y| \, d\mu \leqslant M$.

4. Let $\mathfrak{M}$ be the class of sets of finite measure in S. Let Δ be a binary operation on sets, defined as follows:

$$E \Delta F = (E - F) \cup (F - E) = (E \cup F) - (E \cap F).$$

If E and F are in $\mathfrak{M}$, let $d(E, F) = \mu(E \Delta F)$. Then, with $d(E, F)$ as distance between E and F, $\mathfrak{M}$ is a complete metric space (see Halmos, **2**, pages 168–169).

5. With $\mathfrak{M}$ the metric space defined as in problem 4, show that $L^p(\mu)$ is separable if $\mathfrak{M}$ is separable and if $1 \leqslant p < \infty$. Start with a countable dense subset $\mathfrak{M}_0 \subset \mathfrak{M}$. Then finite linear combinations with rational coefficients, of characteristic functions of members of $\mathfrak{M}_0$, lead to a countable dense subset of $L^p(\mu)$. Use Theorem 7.3–D.

6. Let (T, S, μ) be a measure space such that $T \in S$ and T is a countable union of measurable sets of finite measure. Furthermore, assume that there is a countable collection of members of S such that S coincides with the smallest σ-ring containing this collection. Then the metric space $\mathfrak{M}$ defined in problem 4 is separable. See Halmos, **2**, Theorem B, page 168. Show that the conditions here imposed on (T, S, μ) are fulfilled if T is a Lebesgue measurable subset (not necessarily proper) of Euclidean space of n-dimensions and μ is Lebesgue measure on the measurable subsets of T. Hence $L^p(\mu)$ is separable in this case.

7.4 Continuous Linear Functions on L^p

We work with L^p as defined in § 7.3. To begin with it is convenient to assume $T \in S$ and $\mu(T) < \infty$. This restriction is partially relaxed later.

Theorem 7.4–A. *Suppose* $1 \leqslant p < \infty$. *Suppose* $T \in S$ *and* $\mu(T) < \infty$. *Then there is a one-to-one, linear, norm-preserving correspondence between the normed conjugate space of* L^p *and the space* $L^{p'}$, *such that, if* $x' \in (L^p)'$ *and* $y \in \mathscr{L}^{p'}$ *is a function which represents the element of* $L^{p'}$ *corresponding to* x', *we have*

$$x'(x) = \int xy \, d\mu \tag{7.4–1}$$

for each $x \in \mathscr{L}^p$. *The norm-preserving feature of the correspondence means that*

$$\|x'\| = \left(\int |y|^{p'} \, d\mu\right)^{1/p'} \tag{7.4–2}$$

if $1 < p$, *and*

$$\|x'\| = \sup{}^0 |y(t)| \tag{7.4–3}$$

in case $p = 1$ (*in which case* $p' = \infty$).

PROOF. Let us start with $y \in \mathscr{L}^{p'}$ and define x' by (7.4–1). Then x' is linear and continuous, with $\|x'\| \leqslant \|y\|_{p'}$. This inequality comes from Hölder's inequality if $1 < p < \infty$ and from the fact that $|x(t)y(t)| \leqslant |x(t)| \, \|y\|_\infty$ a.e. in case $p = 1$, $p' = \infty$. The mapping $y \rightarrow x'$ plainly defines a continuous linear mapping of $L^{p'}$ into $(L^p)'$. Also, it is clear by Theorem 7.3–E if $1 < p$, and by problem 2 of § 7.3 if $p = 1$, that if $x'(x) = 0$ for every x, then $\|y\|_{p'} = 0$ and so $y(t) = 0$ a.e. Thus the mapping is one-to-one. It remains to prove that every $x' \in (L^p)'$ is the image of some $y \in L^{p'}$ and that $\|x'\| = \|y\|_{p'}$.

So, suppose $x' \in (L^p)'$. For any $E \in S$ let $\nu(E) = x'(x_E)$, where x_E is the characteristic function of E. Note that $\|x_E\|_p = [\mu(E)]^{1/p}$. Hence

$$|\nu(E)| \leqslant \|x'\| [\mu(E)]^{1/p}. \tag{7.4–4}$$

The set function ν is finitely additive; for, if E and F are disjoint measurable sets and $G = E \cup F$, we have $x_G = x_E + x_F$, whence $\nu(E \cup F) = \nu(E) + \nu(F)$. The countable additivity of ν can now be proved with the aid of (7.4–4). Suppose $\{E_n\}$ is a sequence of pairwise disjoint measurable sets. Then, if

$$E = \bigcup_{k=1}^{\infty} E_k \quad \text{and} \quad F_n = E - \bigcup_{k=1}^{n} E_k,$$

we have

$$\nu(E) = \sum_{k=1}^{n} \nu(E_k) + \nu(F_n).$$

But $\mu(F_n) \rightarrow 0$, since μ is countable additive, and hence $\nu(F_n) \rightarrow 0$, by

(7.4–4). Therefore ν is countably additive. It also follows from (7.4–4) that ν is absolutely continuous. We can now use the Radon-Nikodym theorem to assert the existence of an integrable function y such that

$$\nu(E) = \int_E y \, d\mu$$

for every $E \in S$. Since $\nu(E) = x'(\chi_E)$, it now follows at once that

$$x'(x) = \int xy \, d\mu \tag{7.4–5}$$

for any simple function x. We wish to show next that this same formula holds for any bounded measurable function. If x is bounded and measurable (and hence in $\mathscr{L}^p$), Theorem 7.3–D shows that there exists a uniformly bounded sequence $\{x_n\}$ of simple functions such that $\|x_n - x\|_p \to 0$. The proof of Theorem 7.3–D shows that for the case now before us, in which x is bounded and $\mu(T) < \infty$, we can also require of the sequence $\{x_n\}$ that $x_n(t) \to x(t)$ for each t. It then follows by Lebesgue's theorem of dominated convergence that

$$\int x_n y \, d\mu \to \int xy \, d\mu.$$

Thus, since $x'(x_n) = \int x_n y \, d\mu$, we see that (7.4–5) holds for every bounded measurable x. Since $|x'(x)| \leqslant \|x'\| \, \|x\|_p$, the converse of Hölder's inequality tells us that $y \in \mathscr{L}^{p'}$ and $\|y\|_{p'} \leqslant \|x'\|$ (Theorem 7.3–E if $1 < p$, problem 2, § 7.3 if $p = 1$). Finally, since the elements of L^p corresponding to bounded x are dense in L^p (Lemma 7.3–C), and since $\int xy \, d\mu$ defines a continuous linear functional on *all* of L^p (now that we know $y \in \mathscr{L}^{p'}$), it is clear that (7.3–5) is valid for *all* $x \in \mathscr{L}^p$. We showed that $\|x'\| \leqslant \|y\|_{p'}$ at the outset. Hence $\|x'\| = \|y\|_{p'}$, and the proof is complete.

We now relax the condition $\mu(T) < \infty$.

Theorem 7.4–B. *Theorem* 7.4–A *remains true as stated if the hypothesis that* $\mu(T) < \infty$ *is replaced by the hypothesis that* T *is the union of a countable family of measurable sets of finite measure.*

PROOF. Let G be any fixed measurable set of finite measure. Those elements $E \in S$ such that $E \subset G$ form a σ-ring containing G, and, if we restrict μ to these sets E, it is a measure on this new σ-ring. Let us write $\mathscr{L}^p(T)$ and $\mathscr{L}^p(G)$ to distinguish the function classes which arise when we talk about functions defined over T and G respectively, using the measure μ and its restriction as just described. If g is a function defined on G, let us denote by g^e the function defined on T by

$$g^e(t) = \begin{cases} g(t) & \text{on } G \\ 0 & \text{on } T - G. \end{cases}$$

Evidently $g \in \mathscr{L}^p(G)$ implies $g^e \in \mathscr{L}^p(T)$ and

$$\int_G |g|^p \, d\mu = \int_T |g^e|^p \, d\mu.$$

Let us now suppose that x' is a continuous linear functional on $L^p(T)$. If $g \in \mathscr{L}^p(G)$, let us define $f(g) = x'(g^e)$. It is easily verified that f is a continuous linear functional on $L^p(G)$ and that, as such, its norm satisfies the inequality $\|f\| \leqslant \|x'\|$. By Theorem 7.4–A there exists a function $y \in \mathscr{L}^{p'}(G)$ such that

$$f(g) = \int_G gy \, d\mu$$

for each $g \in \mathscr{L}^p(G)$. Moreover, $\|y\|_{p'} = \|f\| \leqslant \|x'\|$.

Now let us write $T = \bigcup_{n=1}^{\infty} E_n$, where E_n has finite measure. Let $G_n = E_1 \cup \cdots \cup E_n$. We apply the foregoing reasoning to each G_n and obtain corresponding functionals f_n and functions y_n. For any fixed n suppose $g \in \mathscr{L}^p(G_n)$, and let h be the restriction of g^e to G_{n+1}. Then $g^e = h^e$, and we have

$$\int_{G_n} gy_n \, d\mu = x'(g^e) = x'(h^e) = \int_{G_{n+1}} hy_{n+1} \, d\mu = \int_{G_n} gy_{n+1} \, d\mu.$$

The last equality comes from the fact that $h(t) = 0$ on $G_{n+1} - G_n$. We thus have

$$\int_{G_n} g(y_n - y_{n+1}) \, d\mu = 0$$

for each $g \in \mathscr{L}^p(G_n)$. From this we conclude by Theorem 7.4–A that $y_n(t) - y_{n+1}(t) = 0$ a.e. on G_n. Hence, if for each t we define $y(t) = y_m(t)$, where m is the smallest integer for which $t \in G_m$, we see that $y(t) = y_n(t)$ a.e. on G_n for each n. We can conclude that $y \in \mathscr{L}^{p'}(T)$ and $\|y\|_{p'} \leqslant \|x'\|$. If $p' = \infty$ ($p = 1$), this is immediate, for we know that

$$\sup_{t \in G_n}{}^0 |y_n(t)| = \|f_n\| \leqslant \|x'\|$$

for each n, whence $\sup_{t \in G}^0 |y(t)| \leqslant \|x'\|$. If $p' < \infty$ ($1 < p$), we have

$$\int_{G_n} |y|^{p'} \, d\mu = \int_{G_n} |y_n|^{p'} \, d\mu = \|f_n\|^{p'} \leqslant \|x'\|^{p'},$$

and the desired result follows by application of Fatou's lemma to the sequence $v_n|y|^{p'}$ (where v_n is the characteristic function of G_n).

Finally, suppose $x \in \mathscr{L}^p(T)$, and let g_n be the restriction of x to G_n. Then $g_n{}^e = xv_n$, and so

$$x'(xv_n) = f_n(g_n) = \int_{G_n} g_n y_n \, d\mu = \int xv_n y \, d\mu.$$

But

$$\int |xv_n - x|^p \, d\mu = \int_{T-G_n} |x|^p \, d\mu \to 0,$$

as a result of the fact that the set function $\int_E |x|^p \, d\mu$ is countably additive. It then follows that $x'(xv_n) \to x'(x)$, since x' is continuous. It also follows, by Hölder's inequality, that

$$\int xv_n y \, d\mu \to \int xy \, d\mu.$$

Thus we find that $x'(x) = \int xy \, d\mu$ for every $x \in \mathscr{L}^p(T)$.

We have now shown that each x' determines a $y \in \mathscr{L}^{p'}$ with $\|y\|_{p'} \leqslant \|x'\|$. That each such y determines an x' with $\|x'\| \leqslant \|y\|_{p'}$ and that the correspondence between x' and y is one-to-one are shown as in the proof of Theorem 7.4–A. This finishes the argument for Theorem 7.4–B.

For a discussion of the L^p spaces and of the representation of continuous linear functionals on them, using the Daniell approach to the theory of integration, see Loomis, **1**, Chapter III, especially pages 41–43. If $1 < p < \infty$, but not if $p = 1$, Theorem 7.4–A can be extended to the case in which $(T, \boldsymbol{S}, \mu)$ is any measure space. The key fact is that for any $x \in L(\mu)$ the set $N(x)$ is contained in a countable union of measurable sets of finite measure (Halmos, **2**, Theorem F, page 105).

PROBLEMS

1. Under the conditions on T stated in Theorem 7.4–B, L^p is norm reflexive if $1 < p < \infty$.

2. Discuss the possibility of $L(\mu)$ being norm reflexive.

7.41 Complex L^p Spaces

If $(T, \boldsymbol{S}, \mu)$ is a measure space, we can define complex L^p spaces just as we did in the real case, except that we consider functions with complex values. The results developed in §§ 7.3, 7.4 can all be carried over to the complex case. In some parts of the work the complex case can be treated by separation of the functions into real and imaginary parts; in some instances the best procedure seems to be to imitate the proofs of the real case, making a few necessary modifications here and there (such as writing sgn $\bar{z}$ instead of sgn z in certain places).

7.5 Measures in Locally Compact Hausdorff Spaces

In this section we sketch the subject of measures in locally compact Hausdorff spaces, primarily with a view to indicating the fundamental connection between regular Borel measures and positive linear functionals. Our interest is not to develop the subject systematically, with proofs, but to outline the structure of what we need to know in order to obtain a representation theorem for continuous linear functionals on a certain kind of space of continuous functions.

Throughout this section we let T denote a locally compact Hausdorff space. Let $\boldsymbol{S}$ be the σ-ring of Borel sets in T (defined in § 7.1). A real-valued function x defined on T is called *Borel-measurable* if it is measurable with respect to $\boldsymbol{S}$ in the sense defined in § 7.1. It can be shown, in particular, that x is Borel-measurable if it is continuous and if the set $N(x) = \{t : x(t) \neq 0\}$ is contained in a countable union of compact sets.

Definition. A measure μ on $\boldsymbol{S}$ is called a *Borel measure* if $\mu(C) < \infty$ for each compact set C. The measure is called *regular* if for every Borel set E we have

a. $\mu(E) = \sup_C \mu(C)$ (C compact, $C \subset E$).

b. $\mu(E) = \inf_U \mu(U)$ ($E \subset U$, U an open Borel set).

An open set may fail to be a Borel set; however, an open set *is* a Borel set if it is contained in a countable union of compact sets.

We use the following notations:

$\boldsymbol{C}$: the family of compact sets in T.

$\boldsymbol{U}$: the family of open Borel sets in T.

$\boldsymbol{R}$: the family of sets such that each is contained in a countable union of compact sets.

Note that $\boldsymbol{R}$ is a σ-ring and that every subset of a member of $\boldsymbol{R}$ is itself a member of $\boldsymbol{R}$.

If T happens to be compact, all open sets are Borel sets, and $\boldsymbol{R}$ is the class of all subsets of T.

If x is a function defined on T, the closure of the set $N(x)$ is called the *support* of x.

Definition. We denote by $C_\infty(T)$ the class of all real-valued continuous functions x defined on T, such that the support of x is compact. We denote by P the class of those $x \in C_\infty(T)$ such that $x(t) \geqslant 0$ for each $t \in T$.

Evidently $C_\infty(T)$ is a linear space. It is a normed linear space with norm defined by

$$\|x\| = \sup_{t \in T} |x(t)|.$$

The finiteness of $\|x\|$ stems from the fact that x has compact support. If T is compact, then $C_\infty(T)$ is the same as the space $C(T)$ defined in Example 4, § 3.11. The space $C_\infty(T)$ is complete if T is compact.

We write $x \geqslant y$ if $x - y \in P$. Thus $x \geqslant 0$ and $x \in P$ mean the same thing.

Definition. If x' is a linear functional on $C_\infty(T)$ such that $x'(x) \geqslant 0$ whenever $x \geqslant 0$, we say that x' is *positive*. We denote by P' the class of such x'.

If $x \in C_\infty(T)$ and μ is a Borel measure, then x is integrable with respect to μ. For, if C is the support of x and $M = \sup_{t \in C} |x(t)|$, it is clear that $|x(t)| \leqslant M\chi_C(t)$. But the characteristic function χ_C is an integrable simple function; hence x is integrable (see the end of § 7.1).

If μ is a Borel measure and if we define x' on $C_\infty(T)$ by

$$x'(x) = \int x \, d\mu, \tag{7.5–1}$$

it is evident that $x' \in P'$. The rest of this section is devoted to a discussion of this formula. We indicate how it is established that, if we confine attention to *regular* Borel measures, there is a one-to-one correspondence between P' and the class of regular Borel measures, with corresponding elements x' and μ related by (7.5–1). Our exposition is in outline only, with proofs omitted. For details we refer the reader to Halmos, **2**, Chapter X.

Let us begin with an $x' \in P'$. For each $C \in \boldsymbol{C}$ let

$$\lambda(C) = \inf x'(x), \tag{7.5–2}$$

the infimum being taken over all $x \in P$ such that $x(t) \geqslant 1$ on C. The fact that functions x of this kind do exist is shown by Theorem 2.31–B. It can then be shown that this function λ is what is called a *content*. That is, (*a*) $0 \leqslant \lambda(C)$; (*b*) $\lambda(C_1) \leqslant \lambda(C_2)$ if $C_1 \subset C_2$; (*c*) $\lambda(C_1 \cup C_2) \leqslant \lambda(C_1) + \lambda(C_2)$; (*d*) $\lambda(C_1 \cup C_2) = \lambda(C_1) + \lambda(C_2)$ if $C_1 \cap C_2 = \emptyset$. Theorem 2.31–A is used in the proof of (*d*).

Now, there is a procedure by which a content induces a regular Borel measure. The steps in this procedure are as follows: First define $\lambda_*(U)$ for each $U \in \boldsymbol{U}$ by

$$\lambda_*(U) = \sup_C \lambda(C) \qquad (C \text{ compact, } C \subset U). \tag{7.5–3}$$

Then define $\mu^*(E)$ for each $E \in \boldsymbol{R}$ by

$$\mu^*(E) = \inf \lambda_*(U) \qquad (U \in \boldsymbol{U},\ E \subset U). \tag{7.5–4}$$

The function μ^* turns out to be an *outer measure*. That is, (*a*) $\mu^*(E) \geqslant 0$; (*b*) $\mu^*(\emptyset) = 0$; (*c*) $\mu^*(E) \leqslant \mu^*(F)$ if $E \subset F$; (*d*) $\mu^*\left(\bigcup_{n=1}^{\infty} E_n\right) \leqslant \sum_{n=1}^{\infty} \mu^*(E_n)$. From the outer measure μ^* a measure is constructed in the standard way. A set $E \in \boldsymbol{R}$ is called μ^*-measurable (hereafter just measurable) if

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E')$$

for each $A \in \boldsymbol{R}$. The measurable sets form a σ-ring, and on this σ-ring μ^* is a measure which we denote by μ. All sets of outer measure 0 are measurable. Every Borel set is measurable, and μ is a regular Borel measure on the σ-ring of Borel sets. This measure μ, obtained by starting from (7.5–2) is called the regular Borel measure induced by x'. One of the important facts is that

$$\mu(C) = \lambda(C) \tag{7.5–5}$$

for every compact C.

The next major step is to show that, if μ is induced by x', the formula (7.5–1) is valid for each $x \in C_\infty(T)$. Finally, it must be shown that, if we *define* x' by (7.5–1), where μ is a given regular Borel measure and if ν is the regular Borel measure induced by x', then $\nu = \mu$. For this uniqueness proof the regularity of both μ and ν is needed.

An important feature of the correspondence between x' and μ is that, if y' and ν also correspond and if $y' - x' \in P'$, then $\mu(E) \leqslant \nu(E)$ for each $E \in \boldsymbol{S}$. This is clear from (7.5–2), (7.5–5), and the regularity of μ and ν.

7.51 Signed and Complex Borel Measures

As in § 7.5 we consider a locally compact Hausdorff space and the σ-ring $\boldsymbol{S}$ of Borel sets in T.

Definition. By a signed Borel measure on $\boldsymbol{S}$ we mean a signed measure μ (defined in § 7.2) such that $|\mu(C)| < \infty$ if C is compact. This is equivalent to requiring that μ^+ and μ^- both be Borel measures. A complex measure such that its real and imaginary parts are signed Borel measures is called a complex Borel measure.

We shall need the concept of regularity for signed and complex Borel measures. One way to extend this concept would be simply to say that a signed Borel measure μ is regular if μ^+ and μ^- are regular and that a complex Borel measure is regular if its real and imaginary parts are regular. However, we prefer to confine our attention to finite measures and to give the definition in a different way.

Definition. Let μ be a finite complex or signed Borel measure. Then μ is said to be regular if, corresponding to each Borel set E and each $\epsilon > 0$, there exists a compact set C and an open Borel set U such that $C \subset E \subset U$ and such that

$$|\mu(E) - \mu(F)| < \epsilon \tag{7.51–1}$$

whenever F is a Borel set such that $C \subset F \subset U$.

If μ is nonnegative, this definition of regularity is equivalent to the definition given in § 7.5; this is readily seen from the fact that μ is monotone, so that $\mu(C) \leqslant \mu(F) \leqslant \mu(U)$.

The sum of finite regular real or complex Borel measures is again such a measure. For, if $\mu = \mu_1 + \mu_2$ and if we choose C_k and U_k so that (7.51–1) is satisfied for μ_k $(k = 1, 2)$, we can define $C = C_1 \cup C_2$, $U = U_1 \cap U_2$, and μ will then turn out to satisfy (7.51–1) with 2ϵ in place of ϵ. Likewise, regularity is preserved if we multiply μ by a scalar. Using the norm defined by (7.2–5), we can form a normed linear space of real or complex finite regular Borel measures. This space is complete. The completeness will appear later as a by-product of a representation theorem for linear functionals (see Theorem 7.7–G).

Theorem 7.51–A. *If μ is a finite regular signed Borel measure and $\mu = \mu^+ - \mu^-$ is its Jordan decomposition, then μ^+, u^-, and $|\mu|$ are finite regular Borel measures.*

PROOF. It is enough to show that μ^+ is regular, since $\mu^- = (-\mu)^+$, and $-\mu$ is regular if μ is. The regularity of $|\mu|$ follows, since $|\mu| = \mu^+ + \mu^-$. Now let E be a Borel set, and suppose $\epsilon > 0$. Choose a compact set C and an open Borel set U such that $C \subset E \subset U$ and such that $|\mu(E) - \mu(F)| < \epsilon/2$ if $C \subset F \subset U$. Let G be any Borel set contained in $U - C$. Let $A = E \cup (G - E)$, $B = E - G$. Then $C \subset A \subset U$, $C \subset B \subset U$, and so

$$|\mu(E) - \mu(A)| < \epsilon/2, \qquad |\mu(E) - \mu(B)| < \epsilon/2.$$

But

$$\mu(G) = \mu(G \cap E) + \mu(G - E),$$
$$\mu(E) = \mu(G \cap E) + \mu(B),$$
$$\mu(A) = \mu(E) + \mu(G - E),$$

so

$$(7.51\text{–}2) \qquad |\mu(G)| \leqslant |\mu(E) - \mu(B)| + |\mu(A) - \mu(E)| < \epsilon.$$

It now follows from (7.2–1) that

$$\mu^+(U - C) = \sup_{G \subset U - C} \mu(G) \leqslant \epsilon.$$

Now $E = C \cup (E - C)$ and $\mu^+(E - C) \leqslant \mu^+(U - C) \leqslant \epsilon$, so

$$\mu^+(C) = \mu^+(E) - \mu^+(E - C) \geqslant \mu^+(E) - \epsilon.$$

Likewise

$$\mu^+(U) = \mu^+(E) + \mu^+(U - E) \leqslant \mu^+(E) + \mu^+(U - C) < \mu^+(E) + \epsilon.$$

These results show that μ^+ is regular.

Theorem 7.51–B. *If μ is a finite regular complex Borel measure, the total variation $|\mu|$ is a finite regular Borel measure. The real and imaginary parts of μ are finite regular signed Borel measures.*

PROOF. First we show that, for any Borel set E,

$$(7.51\text{–}3) \qquad |\mu|(E) \leqslant 4 \sup_{F \subset E} |\mu(F)|.$$

Let $\mu = \mu_1 + i\mu_2$, where μ_1 and μ_2 have real values. Denote the Jordan decompositions of μ_1 and μ_2 in the usual way. Now, for any Borel set E,

$$\mu_1{}^+(E) = \sup_{F \subset E} \mu_1(F) \leqslant \sup_{F \subset E} |\mu(F)|,$$

$$\mu_1{}^-(E) = \sup_{F \subset E} \{- \mu_1(F)\} \leqslant \sup_{F \subset E} |\mu(F)|,$$

with corresponding results for $\mu_2{}^+$ and $\mu_2{}^-$. Thus, if $E_1, \cdots, E_n$ are pairwise disjoint Borel sets contained in E, and $G = E_1 \cup \cdots \cup E_n$,

$$\begin{aligned} \sum_{k=1}^{n} |\mu(E_k)| &\leqslant \sum_{k=1}^{n} \mu_1{}^+(E_k) + \sum_{k=1}^{n} \mu_1{}^-(E_k) + \sum_{k=1}^{n} \mu_2{}^+(E_k) + \sum_{k=1}^{n} \mu_2{}^-(E_k) \\ &= \mu_1{}^+(G) + \mu_1{}^-(G) + \mu_2{}^+(G) + \mu_2{}^-(G) \\ &\leqslant \mu_1{}^+(E) + \mu_1{}^-(E) + \mu_2{}^+(E) + \mu_2{}^-(E) \leqq 4 \sup_{F \subset E} |\mu(F)|. \end{aligned}$$

In view of the definition of $|\mu|$ [see (7.2–4)], this proves (7.51–3).

Now, to prove that $|\mu|$ is regular, we proceed as in the proof of Theorem 7.51–A as far as (7.51–2), and it then follows from (7.51–3) that $|\mu|(U - C) \leqslant 4\epsilon$. The rest of the proof is then as before, with $|\mu|$ in place of μ^+.

The fact that the real and imaginary parts of μ are regular is seen at once from the defining condition (7.51–1).

7.6 Vector Lattices

In § 7.7, in the discussion of continuous linear functionals on $C_\infty(T)$, we shall find it convenient to use notions which find their general expression in the context of vector lattices. Hence this section is devoted to a brief discussion of vector lattices.

Let A be a partially ordered set with elements $a, b, \cdots$ (see § 1.7), and let the order relation be denoted by $\leqslant$ instead of $\prec$. If $a \leqslant c$ and $b \leqslant c$, we call c an upper bound for a and b. If furthermore $c \leqslant d$ whenever d is an upper bound for a and b, we call c the *least upper bound* of a and b, and write $c = \sup(a, b)$. This element of A is unique, if it exists. In a similar way we define the greatest lower bound of a and b, and denote it by $\inf(a, b)$. If $\sup(a, b)$ and $\inf(a, b)$ exist for every pair a, b in A, A is called a *lattice.*

Definition. By a partially ordered linear space we mean a linear space A which is partially ordered by a relation $\leqslant$ in such a way that the two following conditions are satisfied:

$a \leqslant b$ implies $a + c \leqslant b + c$ for every c.
$a \leqslant b$ implies $\alpha a \leqslant \alpha b$ for every $\alpha > 0$.

If A is moreover a lattice we call it a *vector lattice.*

For each element in a vector lattice we define

$$a^+ = \sup(a, 0), \qquad a^- = -\inf(a, 0). \tag{7.6–1}$$

Theorem 7.6–A. *In a vector lattice A we always have*

$$a = a^+ - a^-, \tag{7.6–2}$$

$$\inf(a^+, a^-) = 0. \tag{7.6–3}$$

PROOF. To prove (7.6–2) we show that $a - a^+ = \inf(a, 0)$. We know that $0 \leqslant a^+$ and $a \leqslant a^+$, by the definition of a^+. Equivalently, then, $a - a^+ \leqslant a$ and $a - a^+ \leqslant 0$. To complete the proof we suppose that $b \leqslant a$ and $b \leqslant 0$ and show that $b \leqslant a - a^+$; it will then follow that $a - a^+ = \inf(a, 0)$. Now $b \leqslant a$ and $b \leqslant 0$ are equivalent to $0 \leqslant a - b$ and $a \leqslant a - b$. Hence, by the definition of a^+, we have $a^+ \leqslant a - b$, or $b \leqslant a - a^+$. This finishes the proof of (7.6–2). The representation of a by (7.6–2) is called the Jordan decomposition of a.

To prove (7.6–3) we note first that $0 \leqslant a^+$ and $0 \leqslant a^-$. Hence (7.6–3) will be proved if we show that $c \leqslant a^+$ and $c \leqslant a^-$ together imply $c \leqslant 0$. We rewrite $c \leqslant a^+$ and $c \leqslant a^-$ in the forms $0 \leqslant a^+ - c$, $a \leqslant a^+ - c$.

It then follows that $a^+ \leqslant a^+ - c$, or $c \leqslant 0$, by the definition of a^+. This completes the proof of Theorem 7.6–A.

It is useful to know the following formulas, which are valid in any vector lattice:

$$\inf(a, b) = -\sup(-a, -b), \tag{7.6–4}$$

$$\sup(a, b) + \inf(a, b) = a + b, \tag{7.6–5}$$

$$\sup(a, b) = a + \sup(b - a, 0). \tag{7.6–6}$$

We omit the proofs, which are much like the proofs of (7.6–2) and (7.6–3).

If A is a partially ordered linear space such that $\sup(a, 0)$ exists for every $a \in A$, then A is a vector lattice. For, it is then easy to show that $\sup(a, b)$ exists for every pair (a, b), being given by (7.6–6); and $\inf(a, b)$ exists, being given by (7.6–4). These remarks are useful when it comes to verifying that we have a vector lattice in various concrete situations.

The space $C_\infty(T)$ is a vector lattice, with the partial ordering $x \leqslant y$ defined to mean that $x(t) \leqslant y(t)$ for every $t \in T$. In this case $z = \sup(x, y)$ is the function defined by $z(t) = \max\{x(t), y(t)\}$. Then $x^+(t) = \max[x(t), 0]$ and $x^-(t) = -\min[x(t), 0]$.

The finite signed measures on a fixed σ-ring form a vector lattice, the partial ordering $\mu \leqslant \nu$ being understood to mean $\mu(E) \leqslant \nu(E)$ for every E in the ring. The Jordan decomposition $\mu = \mu^+ - \mu^-$ described in § 7.2 is what (7.6–2) becomes for this case. The finite regular signed Borel measures on the Borel sets of a locally compact Hausdorff space also form a vector lattice (see Theorem 7.51–A).

In § 7.7 we shall see that the continuous linear functionals defined on $C_\infty(T)$ also form a vector lattice. Here the partial ordering $x' \leqslant y'$ means that $x'(x) \leqslant y'(x)$ whenever $x \geqslant 0$.

7.7 Linear Functionals on $C_\infty(T)$

In this section we use the notations of § 7.5. Our principal aim is to show that we can identify the normed conjugate of the real space $C_\infty(T)$ with the Banach space of all finite regular signed Borel measures on S. This identification is actually, of course, a congruence of the two Banach spaces in question. The Banach spaces are vector lattices, and the linear correspondence is order preserving.

In the course of our work we go somewhat further than a study of the continuous linear functionals on $C_\infty(T)$; we examine linear functionals which may not be continuous, but which have a property which we call *relative boundedness.* It turns out that these linear functionals are precisely those which can be expressed as the difference of two positive linear functionals (as defined in § 7.5). If T is compact, relative boundedness is the same as continuity. A relatively bounded linear functional is called a Radon measure by some authors; see Bourbaki, **5**.

Definition. A linear functional x' defined on $C_\infty(T)$ is said to be relatively bounded if to each compact subset T_0 of T there corresponds a nonnegative number $m(T_0)$ such that

$$|x'(x)| \leqslant m(T_0)\|x\| \tag{7.7–1}$$

whenever $x \in C_\infty(T)$ and the support of x is contained in T_0.

It is easily seen that the relatively bounded linear functionals form a linear space.

Theorem 7.7–A. *A positive linear functional on $C_\infty(T)$ is relatively bounded.*

PROOF. Given a positive x' and a compact set T_0, we choose a neighborhood U of T_0 having compact closure and a function $y \in C_\infty(T)$ such that $y(t) = 1$ on T_0, $y(t) = 0$ on U', and $0 \leqslant y(t) \leqslant 1$ for all t (Theorem 2.31–B). If the support of x is contained in T_0, it is clear that $-\|x\|y(t) \leqslant x(t) \leqslant \|x\|y(t)$ and, therefore, that $-\|x\|x'(y) \leqslant x'(x) \leqslant \|x\|x'(y)$, since x' is positive. We can then choose $m(T_0) = |x'(y)|$ to show that x' is relatively bounded.

Our next main objective is to show that every relatively bounded linear functional is the difference of two positive linear functionals. For this we need some preliminary results.

Lemma 7.7–B. *If f is a real-valued function on $C_\infty(T)$ such that $f(x + y) = f(x) + f(y)$ for every x and y and if $f(x) \geqslant 0$ whenever $x \geqslant 0$, then $f(\alpha x) = \alpha f(x)$ for each real α.*

PROOF. This theorem is actually valid with any vector lattice in place of $C_\infty(T)$, as the proof will show. From $f(0) + f(0) = f(0)$ follows $f(0) = 0$. Then $f(x) + f(-x) = f(0) = 0$, so $f(-x) = -f(x)$. Hence it suffices to prove $f(\alpha x) = \alpha f(x)$ for $\alpha > 0$. It also suffices to show this for $x \geqslant 0$, since every x can be written as $x = x^+ - x^-$ with $x^+ \geqslant 0$, $x^- \geqslant 0$. Now it is easy to deduce that $f(\alpha x) = \alpha f(x)$ if α is positive and rational; we omit the details. It is for the case of irrational α that we need to know that $f(x) \geqslant 0$ if $x \geqslant 0$. Let $\{\alpha_n\}$ and $\{\beta_n\}$ be sequences of positive rationals converging to α such that $\alpha_n < \alpha < \beta_n$. Then, if $x \geqslant 0$, $\alpha_n x \leqslant \alpha x \leqslant \beta_n x$ and $\alpha_n f(x) = f(\alpha_n x) \leqslant f(\alpha x) \leqslant f(\beta_n x) = \beta_n f(x)$. Letting $n \to \infty$, we obtain $\alpha f(x) \leqslant f(\alpha x) \leqslant \alpha f(x)$. This completes the proof.

Lemma 7.7–C. *If f is a real-valued function which is defined, with nonnegative values, for each $x \in C_\infty(T)$ such that $x \geqslant 0$, and if $f(x + y) = f(x) + f(y)$ for each x and y of this kind, there is a unique linear functional F defined on $C_\infty(T)$ such that $F(x) = f(x)$* if $x \geqslant 0$.

PROOF. This result also holds for any vector lattice, not merely on $C_\infty(T)$. Each x can be written in the form $x = x_1 - x_2$ with $x_1 \geqslant 0$ and $x_2 \geqslant 0$. If we have another such representation $x = y_1 - y_2$, then $x_1 + y_2 = x_2 + y_1$, whence it easily follows that $f(x_1) - f(x_2) = f(y_1) - f(y_2)$. Consequently, if we define $F(x) = f(x_1) - f(x_2)$, the value of $F(x)$ depends only on x and not on the way in which x is decomposed. (We note, incidentally, that this is the only value that $F(x)$ could have if it is to be linear and an extension of f. Hence the uniqueness part of the assertion in the theorem.) It is easily verified that $F(x + y) = F(x) + F(y)$, and $F(\alpha x) = \alpha F(x)$ follows by Lemma 7.7–B. Thus the proof is finished.

Theorem 7.7–D. *The relatively bounded linear functionals on $C_\infty(T)$ form a vector lattice. If x' is such a functional, then the Jordan decomposition of x' is*

$$x' = (x')^+ - (x')^-, \tag{7.7–2}$$

where $(x')^+$ and $(x')^-$ are defined as follows:

$$(x')^+(y) = \sup_{x \in P_y} x'(x), \tag{7.7–3}$$

$$(x')^-(y) = -\inf_{x \in P_y} x'(x), \tag{7.7–4}$$

where $y \geqslant 0$ and P_y is the set of all $x \in C_\infty(T)$ such that $0 \leqslant x \leqslant y$.

PROOF. Observe that $\|x\| \leqslant \|y\|$ if $x \in P_y$. Also, if T_0 is the support of y, the support of x is contained in T_0. Hence, by (7.7–1), we have

$$|x'(x)| \leqslant m(T_0)\|y\| \tag{7.7–5}$$

if $x \in P_y$. At this stage we define $f(y) = \sup_{x \in P_y} x'(x)$ if $y \geqslant 0$. Clearly $f(y)$ is finite, by (7.7–5); and $f(y) \geqslant 0$, for $0 \in P_y$. If $0 \leqslant x_1 \leqslant y_1$ and $0 \leqslant x_2 \leqslant y_2$, we have $0 \leqslant x_1 + x_2 \leqslant y_1 + y_2$ and therefore $x'(x_1) + x'(x_2) = x'(x_1 + x_2) \leqslant f(y_1 + y_2)$. It follows that $f(y_1) + f(y_2) \leqslant f(y_1 + y_2)$. To prove the reverse of this inequality, suppose that x, y_1, and y_2 are given, with $x \in P_{y_1+y_2}$. We define

$$x_1(t) = \min \{x(t), y_1(t)\}$$

$$x_2(t) = \max \{0, x(t) - y_1(t)\}.$$

Then x_1 and x_2 are in $C_\infty(T)$ and $x = x_1 + x_2$. It is evident that $0 \leqslant x_1 \leqslant y_1$. It is also true that $0 \leqslant x_2 \leqslant y_2$. For, if $x(t) < y_1(t)$, we have $x_2(t) = 0 \leqslant y_2(t)$, and if $x(t) \geqslant y_1(t)$, we have $x_2(t) = x(t) - y_1(t) \leqslant y_1(t) + y_2(t) - y_1(t) = y_2(t)$. But then $x'(x) = x'(x_1) + x'(x_2) \leqslant f(y_1) + f(y_2)$, and so $f(y_1 + y_2) \leqslant f(y_1) + f(y_2)$. It is now clear that f satisfies the conditions of Lemma 7.7–C and hence that it can be extended in a unique way to give a positive linear functional on $C_\infty(T)$. We denote this extension by $(x')^+$; thus $(x')^+$ is determined by (7.7–3).

It is clear that $C_\infty(T)$ is a partially ordered linear space, and it is seen directly from (7.7–3) that $(x')^+$ is $\sup(x', 0)$. As was remarked in § 7.6, the existence of $\sup(x', 0)$ for each x' insures that we have a vector lattice. As for $(x')^-$, we can either define $(x')^-$ in terms of x' and $(x')^+$ by (7.7–2) and then prove (7.7–4), or we can define $(x')^-$ by (7.7–4), verify that $(x')^-$ is in fact $-\inf(x', 0)$ in the lattice sense, and then obtain (7.7–2) as a special case of (7.6–2). We leave details to the reader.

We define a positive linear functional $|x'|$ by the formula

$$|x'| = (x')^+ + (x')^-. \tag{7.7–6}$$

We must be careful to distinguish $|x'|(x)$ from $|x'(x)|$. Of course we have $(x')^+ = |x'| = x'$ if $x' \geqslant 0$. It may be shown without great difficulty (see problem 2) that

$$|x'|(y) = \sup_{|x| \in P_y} x'(x) \tag{7.7–7}$$

if $y \geqslant 0$. Here $|x|$ denotes the function whose values are $|x(t)|$.

As a consequence of (7.7–7) we see that

$$|x'(x)| \leqslant |x'|(|x|). \tag{7.7–8}$$

For, $x'(x) \leqslant |x'|(|x|)$; on combining this with what is obtained when x is replaced by $-x$, we obtain (7.7–8).

Definition. If x' is any relatively bounded linear functional on $C_\infty(T)$, we define

$$\|x'\| = \sup_{\|x\| \leqslant 1} |x'(x)|. \tag{7.7–9}$$

It may of course happen that $\|x'\| = \infty$ (though not if T is compact). The condition $\|x'\| < \infty$ is satisfied if and only if x' is continuous on $C_\infty(T)$ as a normed linear space. The triangular inequality $\|x' + y'\| \leqslant \|x'\| + \|y'\|$ is valid, with the usual conventions about ∞.

Theorem 7.7–E. *The following relations hold:*

$$\|x'\| = \sup_{0 \leqslant y \leqslant 1} |x'|(y), \tag{7.7–10}$$

$$\|\, |x'| \,\| = \|x'\|. \tag{7.7–11}$$

If $x' \geqslant 0$ *and* $y' \geqslant 0$, *then*

$$\|x' + y'\| = \|x'\| + \|y'\|. \tag{7.7–12}$$

PROOF. We remark that the function with value 1 at all points of T does not belong to $C_\infty(T)$ unless T is compact. Nevertheless, it is convenient to write $y \leqslant 1$ if $y(t) \leqslant 1$ for all t. Suppose that $0 \leqslant y \leqslant 1$ and $|x| \leqslant y$. Then $\|x\| \leqslant 1$. By combining (7.7–7) and (7.7–9) we see that $|x'|(y) \leqslant \|x'\|$. Thus $\sup_{0 \leqslant y \leqslant 1} |x'|(y) \leqslant \|x'\|$. On the other hand, $\|x\| \leqslant 1$ implies $|x| \leqslant 1$; hence (7.7–8) implies

$$|x'(x)| \leqslant \sup_{0 \leqslant y \leqslant 1} |x'|(y)$$

if $\|x\| \leqslant 1$, and from this we infer that

$$\|x'\| \leqslant \sup_{0 \leqslant y \leqslant 1} |x'|(y).$$

This proves (7.7–10). The truth of (7.7–11) follows at once from (7.7–10) and the fact that $|y'| = y'$ if $y' = |x'|$.

To prove (7.7–12) it suffices to prove that $\|x' + y'\| \geqslant \|x'\| + \|y'\|$ under the given conditions, for the reverse inequality is always true. If $0 \leqslant x \leqslant 1$ and $0 \leqslant y \leqslant 1$, let $z(t) = \max \{x(t), y(t)\}$. Then $0 \leqslant z \leqslant 1$, $x \leqslant z$, $y \leqslant z$. Since x' and y' are positive, we have $x'(x) \leqslant x'(z)$, $y'(y) \leqslant y'(z)$, and therefore

$$x'(x) + y'(y) \leqslant (x' + y')(z) \leqslant \|x' + y'\|.$$

Since $x' = |x'|$ and $y' = |y'|$, we now infer from (7.7–10) that $\|x'\| + \|y'\| \leqslant \|x' + y'\|$. Thus (7.7–12) is proved.

Now suppose that x' is a relatively bounded linear functional on $C_\infty(T)$. By the results described in § 7.5 we know that to the positive linear functional $(x')^+$ there corresponds a unique regular Borel measure, which we denote by μ_+, such that

$$\int x \, d\mu_+ = (x')^+(x) \tag{7.7–13}$$

for each $x \in C_\infty(T)$. Likewise, to $(x')^-$ corresponds a unique regular Borel measure μ_- and a formula corresponding to (7.7–13). Hence x' has the representation

$$x'(x) = \int x \, d\mu_+ - \int x \, d\mu_-, \qquad x \in C_\infty(T). \tag{7.7–14}$$

The representation of $|x'|$ is

$$|x'|(x) = \int x \, d\nu, \tag{7.7–15}$$

where $\nu = \mu_+ + \mu_-$.

It should be noted that we do not at this juncture define a signed measure $\mu = \mu_+ - \mu_-$, for we have no assurance that the situation $\mu_+(E) = \mu_-(E) = +\infty$ cannot arise, and thus the unambiguous definition of μ may be impossible. Presently we shall see that in the case when x' is continuous, $\mu_+ - \mu_-$ is indeed the Jordan decomposition of a finite signed measure μ.

We have the following theorem.

Theorem 7.7–F. *If x' is any relatively bounded linear functional on $C_\infty(T)$ and if ν is the regular Borel measure corresponding to $|x'|$, then*

$$\|x'\| = \sup_{E \in S} \nu(E). \tag{7.7–16}$$

PROOF. Because of (7.7–11) it suffices to give the proof on the assumption that $x' \geqslant 0$. In this case $\nu = \mu_+$ and $\mu_- = 0$. Consider any $y \in C_\infty(T)$ with $0 \leqslant y \leqslant 1$, and let C be the support of y. There exists a neighborhood U of C with compact closure (so that U is a Borel set) and a function $x \in C_\infty(T)$ such that $0 \leqslant x \leqslant 1$, $x(t) = 1$ on C and $x(t) = 0$ on $T - U$. Then $y \leqslant x$ and $x'(y) \leqslant x'(x)$. Also

$$x'(x) = \int x \, d\nu \leqslant \nu(U).$$

Thus, if M denotes the supremum on the right in (7.7–16), we have $x'(y) \leqslant M$, whence $\|x'\| \leqslant M$, by (7.7–10). On the other hand, if C is *any* compact set, there exists $x \in C_\infty(T)$ such that $x(t) = 1$ on C and $0 \leqslant x \leqslant 1$. Then, by (7.5–2) and (7.5–5) we see that $\nu(C) \leqslant x'(x)$; also, of course, $x'(x) \leqslant \|x'\|$. By the regularity of ν it then follows that $\nu(E) \leqslant \|x'\|$ for every $E \in S$. This completes the proof of (7.7–16).

If x' is not merely relatively bounded, but continuous, $(x')^+$ and $(x')^-$ are continuous also. This is clear from (7.7–12) as applied to (7.7–6), in view of (7.7–11). Hence the continuous linear functionals on $C_\infty(T)$ form a vector lattice.

We come now to the representation of the normed conjugate of $C_\infty(T)$.

Theorem 7.7–G. *The normed conjugate of $C_\infty(T)$ is congruent as a Banach space to the Banach space of all finite regular signed Borel measures on S. The correspondence between a continuous linear functional x' and a finite regular signed Borel measure μ is established by the formula*

$$x'(x) = \int x \, d\mu \qquad x \in C_\infty(T). \tag{7.7–17}$$

These Banach spaces are also vector lattices, and the correspondence set up by (7.7–17) is order-preserving, so that it is a lattice isomorphism.

PROOF. Starting with x', we obtain the representation (7.7–14). The norm of x' is given by (7.7–16), and since $\nu = \mu_+ + \mu_-$, it is clear from this that the values of μ_+ and μ_- are always finite, because $\|x'\| < \infty$. Hence we can define $\mu = \mu_+ - \mu_-$, and μ is a finite regular signed Borel measure. At this stage we do not yet know that $\mu_+ = \mu^+$ and $\mu_- = \mu^-$. It is not at once evident that the mapping $x' \to \mu$ is linear, but we shall show that it is. From (7.7–3) it is readily apparent that $(x' + y')^+ \leqslant (x')^+ + (y')^+$. Let us write

$$z' = (x')^+ + (y')^+ - (x' + y')^+,$$

so that $z' \geqslant 0$. From (7.7–2) we see that

$$z' = (x')^- + (y')^- - (x' + y')^-$$

also. Now let μ_+, ν_+, ω_+ denote the regular Borel measures associated with $(x')^+$, $(y')^+$, $(x' + y')^+$, respectively, and similarly for μ_-, $(x')^-$, and so on. The measure associated with z' can be expressed either as $\mu_+ + \nu_+ - \omega_+$ or as $\mu_- + \nu_- - \omega_-$. (The fact that $\mu_+ + \nu_+ - \omega_+$ has nonnegative values follows from the fact that the correspondence between positive linear functionals and regular measures is order-preserving; see the very end of § 7.5). But then

$$\omega_+ - \omega_- = (\mu_+ - \mu_-) + (\nu_+ - \nu_-),$$

and this shows that the mapping $x' \to \mu$ is additive. Next, it is evident from (7.7–3) and (7.7–4) that $(\alpha x')^+ = \alpha(x')^+$ and $(\alpha x')^- = \alpha(x')^-$ if $\alpha \geqslant 0$, while $(\alpha x')^+ = -\alpha(x')^-$ and $(\alpha x')^- = -\alpha(x')^+$ if $\alpha < 0$. In either case we see that if $x' \to \mu$ then $\alpha x' \to \alpha\mu$. Thus the mapping is linear. It is clear from (7.7–14) that $\mu = 0$ implies $x' = 0$. Consequently, it will be established that we have a one-to-one mapping onto the space of all finite regular signed Borel measures if we show that every μ is the image of some x'.

We therefore let μ be given, and let its Jordan decomposition be $\mu^+ - \mu^-$. We define x' by (7.7–17) [see (7.2–6)]. If C is the support of x,

$$|x'(x)| \leqslant \int |x|\, d|\mu| \leqslant \|x\|\, |\mu|(C) \leqslant \|x\|\, \|\mu\|$$

[where $\|\mu\|$ is defined by (7.2–5)]. This shows that x' is continuous. Applying (7.7–14) to x', we have

$$\int x\, d\mu_+ - \int x\, d\mu_- = \int x\, d\mu^+ - \int x\, d\mu^-,$$

and hence also

$$\int x\, d(\mu_+ + \mu^-) = \int x\, d(\mu^+ + \mu_-)$$

for each $x \in C_\infty(T)$. But then, by the uniqueness feature of the correspondence between positive linear functionals and regular Borel measures (described in § 7.5), we infer that $\mu_+ + \mu^- = \mu^+ + \mu_-$, or $\mu = \mu_+ - \mu_-$. Hence μ is the image of x' under the mapping described at the outset of the proof.

The fact that the mapping is order-preserving follows from the fact that it is one-to-one plus the fact that every positive functional is representable with the aid of a nonnegative measure. The mapping then preserves least upper bounds and greatest lower bounds. From this it follows that $\mu_+ = \mu^+$, for μ_+ is the image of $(x')^+ = \sup(x', 0)$, while $\mu^+ = \sup(\mu, 0)$. Likewise $\mu_- = \mu^-$.

Finally, the mapping is norm-preserving. This follows from (7.7–16) and (7.2–5), for we now know that $\mu_+ + \mu_- = |\mu|$. Since the normed conjugate of $C_\infty(T)$ is complete, it follows that the space of finite regular signed Borel measures is also complete.

PROBLEMS

1. Carry out at least one of the suggestions for dealing with $(x')^-$ as indicated at the end of the proof of Theorem 7.7–D.

2. Prove (7.7–7). Suggestions: If $0 \leqslant |x| \leqslant y$, note that $0 \leqslant x^+ \leqslant y$ and $0 \leqslant x^- \leqslant y$. Then $x'(x) \leqslant |x'|(y)$ follows from (7.7–3) and (7.7–4). On the other hand, if $\epsilon > 0$, choose x_1 and x_2 in P_y so that $x'(x_1) > (x')^+(y) - (\epsilon/2)$ and $x'(x_2) < -(x')^-(y) + (\epsilon/2)$. Then consider $x = x_1 - x_2$.

3. Prove that $\|x'\| = \|(x')^+\| + \|(x')^-\|$.

4. Prove that $\|x'\| \leqslant \|y'\|$ if $0 \leqslant x' \leqslant y'$. Use (7.7–12).

5. If T is compact and x' is a continuous linear functional on $C_\infty(T)$ $[\equiv C(T)]$, show that $\|x'\| = |x'|(1)$, where 1 denotes the function of all whose values are unity.

6. If (μ_1, μ_2) and (ν_1, ν_2) are two pairs of regular Borel measures on $\mathbf{S}$, we define these pairs to be equivalent if $\mu_1(E) + \nu_2(E) = \mu_2(E) + \nu_1(E)$ for every $E \in \mathbf{S}$. Show that

$$\int x\,d\mu_1 - \int x\,d\mu_2 \quad \text{and} \quad \int x\,d\nu_1 - \int x\,d\nu_2$$

define the same relatively bounded linear functional on $C_\infty(T)$ if and only if (μ_1, μ_2) and (ν_1, ν_2) are equivalent. Show that there is a one-to-one correspondence between the set of all relatively bounded linear functionals on $C_\infty(T)$ and the set of all equivalence classes of pairs of regular Borel measures.

7. For the purpose of extending some of the results of § 7.7 to the complex case, let $C_\infty(T)$ denote the set of complex-valued continuous functions on T, each with compact support. Denote the corresponding space of real functions by $RC_\infty(T)$. If z' is a relatively bounded complex linear functional on $C_\infty(T)$, if we write $z' = x' + iy'$, where x' and y' have real values, and if we write elements of $C_\infty(T)$ in the form $z = x + iy$ with x and $y \in RC_\infty(T)$, show that

$$z'(z) = x'(x) - y'(y) + iy'(x) + ix'(y).$$

Show that $z'(z)$ is representable in terms of eight real integrals involving four regular Borel measures. Show also that

$$\sup_{\|z\| \leqslant 1} |z'(z)| = \sup_{\|z\| \leqslant 1} |x'(x) - y'(y)|.$$

Then prove that $\|z'\|$, defined by the foregoing supremum, which may be infinite, satisfies the inequalities

$$(\|x'\|^2 + \|y'\|^2)^{\frac{1}{2}} \leqslant \|z'\| \leqslant \|x'\| + \|y'\|.$$

Both extremes are attainable.

8. Continuing along the lines begun in (7), show that the normed conjugate of $C_\infty(T)$, in the complex case, is congruent to the space of finite regular complex Borel measures on S. The principal difficulty is in showing that the one-to-one correspondence is norm preserving. It suffices to show that, if z' and μ correspond by the formula $z'(z) = \int z\, d\mu$, then $\|z'\| \geqslant \|\mu\|$, for the reverse inequality is evident. We outline the argument. With $\epsilon > 0$ given, choose a Borel set E so that $|\mu|(E) > \|\mu\| - (\epsilon/3)$. Then choose pairwise disjoint Borel sets $E_1, \cdots, E_k$ such that $E_k \subset E$ and

$$\sum_{k=1}^{n} |\mu(E_k)| > |\mu|(E) - (\epsilon/3).$$

Now, if $\|z\| \leqslant 1$,

$$\|z'\| \geqslant \left|\int z\, d\mu\right| \geqslant \sum_{k=1}^{n} |\mu(E_k)| - \left|\int z\, d\mu - \sum_{k=1}^{n} |\mu(E_k)|\right|$$

$$\geqslant \|\mu\| - \tfrac{2}{3}\epsilon - \left|\int z\, d\mu - \sum_{k=1}^{n} |\mu(E_k)|\right|.$$

Hence it suffices to be able to choose z so that $\|z\| \leqslant 1$ and

$$\left|\int z\, d\mu - \sum_{k=1}^{n} |\mu(E_k)|\right| < \epsilon/3.$$

For each k let C_k and U_k be, respectively, compact and open Borel sets such that $C_k \subset E_k \subset U_k$ and such that $|\mu(E_k) - \mu(F_k)| < (\epsilon/27n)$ if F_k is a Borel set for which $C_k \subset F_k \subset U_k$. One can then show, just as in the proof of the regularity of $|\mu|$ in Theorem 7.51–B, that $|\mu|(U_k - C_k) \leqslant (8/27n)$. It also follows that

$$\left|\sum_{k=1}^{n} |\mu(C_k)| - \sum_{k=1}^{n} |\mu(E_k)|\right| \leqslant \sum_{k=1}^{n} |\mu(C_k) - \mu(E_k)| < \epsilon/27.$$

For each i and j with $i \neq j (i, j = 1, \cdots, n)$ there exist disjoint neighborhoods V_{ij} and V_{ji} of C_i and C_j respectively, with compact closures. Let V_k be the intersection of U_k and all the V_{kj} for which $j \neq k$. Then $C_k \subset V_k \subset U_k$, V_k is open with compact closure, and $V_k \cap V_l = \emptyset$ if $k \neq l$. Choose $x_k \in RC_\infty(T)$ so that $0 \leqslant x_k \leqslant 1$, $x_k(t) = 1$ on C_k and $x_k(t) = 0$ on $T - V_k$. Let $\alpha_k = \overline{\operatorname{sgn} \mu(C_k)}$ and $z = \sum_{k=1}^{n} \alpha_k x_k$. This element of $C_\infty(T)$ will do what is required for the completion of the proof.

7.8 Finitely Additive Set Functions

In the proof of Theorem 7.4–A we used the fact that $p < \infty$ in proving that the set function ν is countably additive; the fact that ν is finitely additive remains true even in the case $p = \infty$. Finitely additive set functions turn out to be of interest in connection with the study of linear functionals on spaces of bounded or essentially bounded functions.

Definition. Let T be a nonempty set, $\boldsymbol{R}$ a ring of subsets of T, and μ a finitely additive set function on $\boldsymbol{R}$ with real or complex values. If the values of μ on $\boldsymbol{R}$ are bounded, we call μ a *charge* on $\boldsymbol{R}$. If μ is a charge and $T \in \boldsymbol{R}$, we call $(T, \boldsymbol{R}, \mu)$ a *totally charged space.*

This terminology concerning charges is due to Rosenbloom, **1**.

Being finitely additive means that $\mu(E \cup F) = \mu(E) + \mu(F)$ if E and F are disjoint members of $\boldsymbol{R}$. Of course $\boldsymbol{R}$ contains the empty set $\emptyset$, and hence $\mu(\emptyset) = 0$.

A real-valued charge μ has a Jordan decomposition $\mu = \mu^+ - \mu^-$, just like a signed measure (§ 7.2). With a charge μ we associate the total variation $|\mu|$, given by (7.2–4), just as with a signed or complex measure; $|\mu|$ is a nonnegative charge. We define the norm $\|\mu\|$ of a charge μ just as in (7.2–5). The charges on $\boldsymbol{R}$ form a Banach space. If the space is totally charged, $\|\mu\| = |\mu|(T)$.

Charges can be used to represent the normed conjugate of the space $B(T)$ (defined in Example 3, § 3.11). First we must define a type of integral with respect to a charge. This integral is to some extent analogous to the classical Riemann-Stieltjes integral and is defined with the aid of the notion of Moore-Smith convergence. For terminology about directed sets, nets, and Moore-Smith convergence see Kelley, **1**, Chapter 2.

Suppose that $(T, \boldsymbol{R}, \mu)$ is a totally charged space. By a *partition* of T we mean a finite collection $E_1, \cdots, E_n$ of pairwise disjoint nonempty members of $\boldsymbol{R}$ such that $T = E_1 \cup \cdots \cup E_n$. If $\pi_1 = (E_1, \cdots, E_m)$ and $\pi_2 = (F_1, \cdots, F_n)$ are partitions, we define $\pi_1 \leqslant \pi_2$ to mean that each F_j is a subset of some E_k (π_2 is a "refinement" of π_1). This gives a partial ordering of partitions, and every pair of partitions has an upper bound, e.g., the "superposition" of the two partitions. Hence partitions form a directed set. Now suppose that x is a scalar-valued function defined on T. If $\pi = (E_1, \cdots, E_n)$ is a partition, choose points $t_k \in E_k$ and write

$$f(\pi; t_1, \cdots, t_n) = \sum_{k=1}^{n} x(t_k)\mu(E_k). \tag{7.8–1}$$

The (many-valued) function f is a net on the directed set of partitions. If the net converges to a limit I, this limit is called the S-integral of x with respect to μ:

$$I = S\int x \, d\mu.$$

This means, of course, that for every $\epsilon > 0$ there is some partition π_ϵ such that $f(\pi; t_1, \cdots, t_n)$ differs from I by less than ϵ if $\pi_\epsilon \leqslant \pi$.

The function x is called measurable if $x^{-1}(M) \in \boldsymbol{R}$ for each semi-open interval $M = \{t : \alpha < t \leqslant \beta\}$. This is for the real case. In the complex case we use semi-open rectangles. It can then be proved that when $(T, \boldsymbol{R}, \mu)$ is a totally charged space and x is a bounded measurable function on T, then x is S-integrable. The argument depends on a Cauchy condition for the convergence of nets. We give the argument for the complex case.

Choose a semi-open rectangle M whose interior contains all the values of x. Suppose $\epsilon > 0$, and partition M into disjoint semi-open rectangles $M_1, \cdots, M_m$, all of diagonal less than ϵ. Let $F_k = x^{-1}(M_k)$, and let $E_1, \cdots, E_n$ be the distinct nonempty sets among $F_1, \cdots, F_m$. Then $\pi = (E_1, \cdots, E_n)$ is a partition of T. Suppose now that $\pi_1 = (G_1, \cdots)$ and $\pi_2 = (H_1, \cdots)$ are partitions of T such that $\pi \leqslant \pi_1$ and $\pi \leqslant \pi_2$, and choose points $s_i \in G_i$, $t_j \in H_j$. We can write

$$\mu(G_i) = \sum_j \mu(G_i \cap H_j), \quad \sum_i x(s_i)\mu(G_i) = \sum_{i,j} x(s_i)\mu(G_i \cap H_j),$$

and so we see that

$$\sum_i x(s_i)\mu(G_i) - \sum_j x(t_j)\mu(H_j) = \sum_{i,j} \{x(s_i) - x(t_j)\}\mu(G_i \cap H_j). \tag{7.8–2}$$

We need only consider those terms for which $\mu(G_i \cap H_j) \neq 0$, and for such terms G_i and H_j are both contained in the same E_k, which means that $x(s_i)$ and $x(t_j)$ are contained in the same M_l, and hence that $|x(s_i) - x(t_j)| < \epsilon$. Consequently, the absolute value of the expression in (7.8–2) does not exceed

$$\epsilon \sum_{i,j} |\mu(G_i \cap H_j)| \leqslant \epsilon|\mu|(T).$$

This shows that the net satisfies a Cauchy condition and is therefore convergent.

The S-integral evidently satisfies the inequality

$$\left| S \int x \, d\mu \right| \leqslant \sup_{t \in T} |x(t)| \; |\mu|(T). \tag{7.8–3}$$

Now consider the Banach space $B(T)$, where T is any nonempty set. Let $\boldsymbol{R}$ be the ring of *all* subsets of T. Then any member of $B(T)$ is S-integrable with respect to any charge on $\boldsymbol{R}$.

Theorem 7.8–A. *With T and $\boldsymbol{R}$ as just described, the normed conjugate of $B(T)$ is congruent to the space of all charges on $\boldsymbol{R}$, the correspondence between a continuous linear functional x' and its associated charge μ being indicated by the two formulas*

$$x'(x) = S \int x \, d\mu, \qquad x \in B(T), \tag{7.8–4}$$

$$\mu(E) = x'(x_E), \qquad E \in \boldsymbol{R}, \tag{7.8–5}$$

where x_E is the characteristic function of E.

PROOF. Each charge μ defines an $x' \in \{B(T)\}'$ by formula (7.8–4). From (7.8–3) we see that $\|x'\| \leqslant \|\mu\|$. Formula (7.8–5) clearly holds if E is $\emptyset$ or T. For any other E it suffices to observe that, if π_0 is any partition of T and $\pi = (E_1, \cdots, E_n)$ is the superposition of π_0 and the partition $(E, T - E)$, the expression in (7.8–1) is equal to $\mu(E)$. From (7.8–5) we see that $x' = 0$ implies $\mu = 0$.

On the other hand, if we start with x' and define μ by (7.8–5), it is easy to see that μ is a charge. We wish to show that (7.8–4) holds. If $\epsilon > 0$ and $x \in B(T)$, choose $M, M_1, \cdots, M_m$ and $E_1, \cdots, E_n$ as in the earlier proof of integrability of a bounded measurable function. Choose any $t_k \in E_k$ and form the sum (7.8–1). Since $\mu(E) = x'(x_E)$, we see that

$$\sum_k x(t_k)\mu(E_k) = x'\left(\sum_k x(t_k)y_k\right),$$

where y_k is the characteristic function of E_k. On the other hand, if $t \in E_j$,

$$\left| x(t) - \sum_k x(t_k)y_k(t) \right| = |x(t) - x(t_j)|,$$

so that

$$\|x - \textstyle\sum x(t_k)y_k\| \leqslant \epsilon \tag{7.8–6}$$

and therefore

$$\left| x'(x) - \sum_k x(t_k)\mu(E_k) \right| \leqslant \epsilon\|x'\|.$$

This inequality evidently remains valid for any sums based on partitions obtained by further subdivision of the E_k's and so (7.8–4) is true.

Finally, to show that $\|\mu\| \leqslant \|x'\|$ in the context of the foregoing paragraph, suppose $\epsilon > 0$ and choose a partition $(E_1, \cdots, E_n)$ of T such that

$$\sum_{k=1}^{n} |\mu(E_k)| > \|\mu\| - \epsilon;$$

this is possible, by the definition of $\|\mu\|$. Then define x by setting $x(t) = \overline{\operatorname{sgn}}\, \mu(E_k)$ on E_k. Then $\|x\| \leqslant 1$ and

$$x'(x) = S\int x \, d\mu = \sum_k |\mu(E_k)|.$$

Consequently $\|x'\| \geqslant \|\mu\| - \epsilon$. Thus $\|x'\| \geqslant \|\mu\|$. This establishes everything necessary for the proof of Theorem 7.8–A.

7.9 Lebesgue Integrals with Respect to a Charge

Let T be any nonempty set, and let S be a σ-ring of subsets of T such that $T \in S$. Let N be a nonempty proper subclass of S which is closed under the formation of countable unions and which has the property that any subset of a member of N is again a member of N. Evidently N is closed under the formation of differences and intersections and contains the empty set $\emptyset$. One possibility is that N consists solely of the set $\emptyset$. However, the situation we have in mind as a model is that in which S is the class of Lebesgue measurable sets in Euclidean space and N is the class of sets of measure 0.

If x is a real-valued function on T which is measurable with respect to S, we say that x is N-bounded if $\{t: |x(t)| > \alpha\} \in N$ for some $\alpha \geqslant 0$. The smallest α for which this is true is denoted by $\|x\|$ (the existence of this smallest α requires the countable union closure property of N). If x and y are N-bounded, measurable and if $\|x - y\| = 0$, we call them N-equivalent. Just as in the case of L^∞, we construct a Banach space whose elements are equivalence classes of such functions. We permit ourselves the abuse of language whereby we identify x and the equivalence class to which it belongs, and we denote the Banach space here described by $B_N(T)$. If N contains only the set $\emptyset$ and S is the σ-ring of all subsets of T, $B_N(T)$ is just $B(T)$, as previously defined. If μ is a measure on S such that the sets of zero measure form a class N with the required properties, $B_N(T)$ coincides with $L^\infty(\mu)$, as previously defined [problem 1, § 7.3].

Now suppose that μ is a charge on S such that $\mu(E) = 0$ if $E \in N$, and suppose $x \in B_N(T)$. We can define an integral of x with respect to μ by the classical procedure of Lebesgue. Write $\alpha = -\|x\|$, $\beta = \|x\|$, assuming that $\|x\| > 0$. Suppose $\alpha = \alpha_0 < \alpha_1 < \cdots < \alpha_n = \beta$ and let

$$E_0 = \{t:\alpha_0 = x(t)\}, \qquad E_k = \{t:\alpha_{k-1} < x(t) \leqslant \alpha_k\}, \qquad k = 1, \cdots, n.$$

If $\xi_0 = \alpha_0$ and if $\xi_1, \cdots, \xi_n$ are chosen so that $\alpha_{k-1} \leqslant \xi_k \leqslant \alpha_k$, the sums

$$\sum_{k=0}^{n} \xi_k \mu(E_k)$$

converge to a limit as $\max\limits_k |\alpha_k - \alpha_{k-1}| \to 0$, the limit being independent of how the points $\xi_1, \cdots, \xi_n$ are chosen within the specified subintervals. This limit is by definition the L-integral of x with respect to μ; we denote it by

$$L \int x \, d\mu.$$

The fact that the integral exists can be proved much as in the classical Lebesgue case, and we omit details.

The fact that we assumed $\mu(E) = 0$ if $E \in N$ shows that the value of the integral depends only on the equivalence class represented by the particular N-bounded function, not on the function itself. The definition of the L-integral does not require the introduction of the class N, provided we assume that x is bounded and measurable. Nor is it necessary to have S a σ-ring; merely a ring is enough. However, when we introduce N, we assume that S is a σ-ring.

We can also deal with complex-valued functions, either directly or by separating into real and imaginary parts.

It follows from the form of the sums defining the integral that

$$\left| L \int x \, d\mu \right| \leqslant \|x\| \, \|\mu\|.$$

It may be proved much as in the classical Lebesgue theory that the integral is a linear function of x. Linearity in μ is evident, but linearity in x is not as evident here as with the S-integral. Because of the linearity in μ it is enough, in proving the linearity in x, to treat the case in which μ has nonnegative values, for the general case can be reduced to this by the Jordan decomposition of μ. We omit the details. See, for example, Titchmarsh, **1**, or Rosenbloom, **1**.

If x is measurable and bounded, it is easy to prove that

$$L \int x \, d\mu = S \int x \, d\mu. \tag{7.9–1}$$

For, if we use the simple function

$$y = \sum_k x(t_k)y_k$$

related to x as in the construction leading up to (7.8–6), we see that

$$L\int y\,d\mu = \sum_k x(t_k)\mu(E_k),$$

$$\left|L\int x\,d\mu - L\int y\,d\mu\right| = \left|L\int (x - y)\,d\mu\right| \leqslant \|x - y\|\,\|\mu\| \leqslant \epsilon\|\mu\|,$$

$$\left|S\int x\,d\mu - L\int y\,d\mu\right| \leqslant \epsilon\|\mu\|,$$

and (7.9–1) follows from these results.

Theorem 7.9–A. *The normed conjugate of $B_N(T)$ and the subspace of the space of all charges, consisting of those charges μ such that $\mu(E) = 0$ if $E \in N$, are congruent under the correspondence indicated by the formula*

$$x'(x) = L\int x\,d\mu.$$

The argument is much the same as in the proof of Theorem 7.8–A. Note that, when μ is obtained from x' by the formula $\mu(E) = x'(x_E)$, it follows that $\mu(E) = 0$ if $E \in N$, for $\|x_E\| = 0$ in that case.

SUGGESTED READING

Banach, **1**, pp. 59–65.
Birkhoff, **1**, Chapter 15.
Bourbaki, **5**.
Dunford, **3**, pp. 31–41.
Dunford and Schwartz, **1**, Chapters 3, 4.
Fichtenholz and Kantorovitch, **1**.
Halmos, **2**, Chapters 2, 4, 5, 6, 10.
Hildebrandt, **1**.
Loomis, **1**, Chapter 3.
Riesz and Nagy, **1**, parts of Chapters 2, 3.
Rosenbloom, **1**.
Yosida and Hewitt, **1**.
Zaanen, **1**, Chapter 5 and part of Chapter 7.

BIBLIOGRAPHY

N. I. Achieser and I. M. Glasmann

1. *Theorie der linearen Operatoren im Hilbert-Raum*, Akademie Verlag, Berlin, 1954.

W. G. Bade

1. An operational calculus for operators with spectrum in a strip, *Pacific J. Math.*, 3 (1953), 257–290.

S. Banach

1. *Opérations linéaires*, Monografje Matematyczne, Warsaw, 1932.

G. Birkhoff

1. *Lattice Theory*, revised edition, American Mathematical Society Colloquium Publications, vol. 25, New York, 1948.

R. P. Boas

1. Entire functions of exponential type, *Bull. Amer. Math. Soc.*, 48 (1942), 839–849.

N. Bourbaki

Each of the following is in the series, *Éléments de mathématique*, *Actualités Sci. Ind.*, Hermann, Paris. Each title is identified by a serial number.

1. *Théorie des ensembles.* No. 846 (1939).
2. *Topologie générale.* No. 858 (1940).
3. *Algèbre linéaire.* No. 1032 (1947).
4. *Espaces vectoriels topologiques.* No. 1189 (1953) and No. 1229 (1955).
5. *Intégration.* No. 1175 (1952).

R. V. Churchill

1. *Fourier Series and Boundary Value Problems*, McGraw-Hill, New York, 1941.

J. A. Clarkson

1. Uniformly convex spaces, *Trans. Amer. Math. Soc.*, 40 (1936), 396–414.

R. G. Cooke

1. *Linear Operators*, Macmillan, London, 1953.

R. Courant and D. Hilbert

1. *Methoden der mathematischen Physik*, vol. I, Springer, Berlin, 1931.

M. M. Day

1. The spaces L^p with $0<p<1$, *Bull. Amer. Math. Soc.*, 46 (1940), 816–823.

J. Dieudonné

1. La dualité dans les espaces vectorielles topologiques, *Ann. Sci. Ecole Norm. Sup.* (3), 59 (1942), 107–139.
2. Recent developments in the theory of locally convex vector spaces, *Bull. Amer. Math. Soc.*, 59 (1953), 495–512.

J. Dixmier

1. Sur un théorème de Banach, *Duke Math. J.*, 15 (1948), 1057–1071.
2. Sur les bases orthonormales dans les espaces préhilbertiens, *Acta Sci. Math. Szeged*, 15 (1953), 29–30.

N. Dunford

1. Spectral theory I. Convergence to projections, *Trans. Amer. Math. Soc.*, 54 (1943), 185–217.
2. Spectral theory, *Bull. Amer. Math. Soc.*, 49 (1943), 637–651.
3. Spectral theory in abstract spaces and Banach algebras, *Proceedings of the Symposium on Spectral Theory and Differential Problems*, Oklahoma Agricultural and Mechanical College, 1951.
4. Spectral theory II. Resolutions of the identity, *Pacific J. Math.*, 2 (1952), 559–614.
5. Spectral operators, *Pacific J. Math.*, 4 (1954), 321–354.

N. Dunford and B. J. Pettis

1. Linear operators on summable functions, *Trans. Amer. Math. Soc.*, 47 (1940), 323–392.

N. Dunford and J. T. Schwartz

1. *Linear Operators*, vol. I. *General Theory*, Interscience, New York, 1958.

W. F. Eberlein

1. Weak compactness in Banach spaces I, *Proc. Nat. Acad. Sci., U.S.A.*, 33 (1947), 51–53.

G. Fichtenholz and L. Kantorovitch

1. Sur les opérations dans l'espace des fonctions bornées, *Studia Math.*, 5 (1934), 69–98.

M. Fréchet

1. *Les espaces abstraits*, Gauthier-Villars, Paris, 1928.

B. Friedman

1. *Principles and Techniques of Applied Mathematics*, John Wiley, New York, 1956.

L. M. Graves

1. *The Theory of Functions of Real Variables*, 2nd ed., McGraw-Hill, New York, 1956.

C. J. A. Halberg Jr.

1. Spectral theory of linked operators in the spaces l^p, Doctoral dissertation, University of California, Los Angeles, 1955.

D. W. Hall and G. L. Spencer

1. *Elementary Topology*, John Wiley, New York, 1955.

P. R. Halmos

1. *Finite Dimensional Vector Spaces*, Princeton University Press, 1948.
2. *Measure Theory*, D. Van Nostrand, New York, 1950.
3. *Introduction to Hilbert Space and the Theory of Spectral Multiplicity*, Chelsea, New York, 1951.

H. L. Hamburger and M. E. Grimshaw

1. *Linear Transformations in n-dimensional Vector Space*, Cambridge University Press, 1951.

G. Hamel

1. Eine Basis aller Zahlen und die unstetige Lösungen der Funktionalgleichung $f(x + y) = f(x) + f(y)$, *Math. Ann.*, 60 (1905), 459–462.

G. H. Hardy

1. The mean value of the modulus of an analytic function, *Proc. London Math. Soc.*, 14 (1914), 269–277.

G. H. Hardy, J. E. Littlewood, and G. Pólya

1. *Inequalities*, Cambridge University Press, 1934.

E. D. Hellinger

1. Spectra of quadratic forms in infinitely many variables, *Mathematical Monographs* I (1941), Northwestern University Studies.

E. Hellinger and O. Toeplitz

1. *Integralgleichungen und Gleichungen mit unendlichvielen Unbekannten*, Chelsea, New York, 1953. (A reprint of Part II C13 of the *Encyklopädie der mathematischen Wissenschaften.*)

E. Hille

1. *Functional analysis and semi-groups*, American Mathematical Society Colloquium Publications, vol. 31, New York, 1948.

T. H. Hildebrandt

1. On bounded linear transformations, *Trans. Amer. Math. Soc.*, 36 (1934), 868–875.

D. H. Hyers

1. Linear topological spaces, *Bull. Amer. Math. Soc.*, 51 (1945), 1–24.

N. Jacobson

1. *Lectures in abstract algebra*, vol. II, *Linear Algebra*, D. Van Nostrand, New York, 1953.

G. Julia

1. *Introduction mathématique aux théories quantiques* (second part), Gauthier-Villars, Paris, 1938.

S. Kakutani

1. Weak topology and regularity of Banach spaces, *Proc. Imp. Acad. Tokyo*, 15 (1939), 169–173.

J. L. Kelley

1. *General Topology*, D. Van Nostrand, New York, 1955.

O. D. Kellogg

1. *Foundations of Potential Theory*, Springer, Berlin, 1929.

P. D. Lax

1. On the existence of Green's function, *Proc. Amer. Math. Soc.*, 3 (1952), 526–531.

A. E. Livingston

1. The space H^p, $0 < p < 1$, is not normable, *Pacific J. Math.*, 3 (1953), 613–616.

L. A. Ljusternik and W. I. Sobolew

1. *Elemente der Funktionalanalysis*, Akademie Verlag, Berlin, 1955.

L. H. Loomis

1. *An Introduction to Abstract Harmonic Analysis*, D. Van Nostrand, New York, 1953.

E. R. Lorch

1. The spectrum of linear transformations, *Trans. Amer. Math. Soc.*, 52 (1942), 238–248.

H. Löwig

1. Über die Dimension Linearer Räume, *Studia Math.*, 5 (1934), 18–23.

G. W. Mackey

1. On infinite dimensional linear spaces, *Trans. Amer. Math. Soc.*, 57 (1945), 155–207.
2. On convex topological linear spaces, *Trans. Amer. Math. Soc.*, 60 (1946), 519–537.

E. J. McShane

1. *Integration*, Princeton University Press, 1947.

D. Milman

1. On some criteria for the regularity of spaces of the type (B), *C. R.* (*Doklady*) *de l'Acad. Sci. de l'URSS*, 20 (1938), 243–246.

C. Miranda

1. Sul principio di Dirichlet per le funzioni armoniche, *Atti Accad. Naz. Lincei, Rend. Cl. Sci. Fis. Mat. Nat.* (8), 3 (1947), 55–59.

F. J. Murray

1. On complementary manifolds and projections in spaces L_p and l_p, *Trans. Amer. Math. Soc.*, 41 (1937), 138–152.

B. v. Sz. Nagy

1. *Spektraldarstellung linearer Transformationen des Hilbertschen Raumes*, *Ergebnisse der Mathematik und ihrer Grenzgebiete*, Springer, Berlin, 1942.

J. von Neumann

1. *Mathematische Grundlagen der Quantenmechanik*, Springer, Berlin, 1932.
2. On complete topological spaces, *Trans. Amer. Math. Soc.*, 37 (1935), 1–20.

B. J. Pettis

1. A proof that every uniformly convex space is reflexive, *Duke Math. J.*, 5 (1939), 249–253.

R. S. Phillips

1. On linear transformations, *Trans. Amer. Math. Soc.*, 48 (1940), 516–541.

H. Pollard

1. Integral transforms, *Duke Math. J.*, 13 (1946), 307–330.

F. Rellich

1. Spektraltheorie in nichtseparablen Räumen, *Math. Ann.*, 110 (1934), 342–356.

F. Riesz

1. *Les systèmes d'équations lineairés à une infinité d'inconnues*, Gauthier-Villars, Paris, 1913.
2. Über die Randwerte einer analytischen Funktion, *Math. Zeit.*, 18 (1923), 87–95.
3. Sur les valeurs moyennes du module des fonctions harmoniques et des fonctions analytiques, *Acta Sci. Math. Szeged*, 1 (1922), 27–32.

F. Riesz and B. v. Sz. Nagy

1. *Leçons d'analyse fonctionelle*, 3rd ed. Gauthier-Villars, Paris, 1955.

M. Riesz

1. Sur les maxima des formes bilinéaires et sur les fonctionelles linéaires, *Acta Math.*, 49 (1927), 465–497.

P. C. Rosenbloom

1. Quelques classes de problèmes extrémaux, *Bull. Sci. Math. France*, 79 (1951), 1–58, and 80 (1952), 183–215.

S. Saks

1. *Theory of the Integral*, 2nd ed., Monografje Matematyczne, Warsaw-Lwów, 1937.

W. Schmeidler

1. *Lineare Operatoren in Hilbertschen Raum*, Teubner, Stuttgart, 1954.

I. Schur

1. Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlichvielen Veränderlichen, *J. Reine Angew. Math.*, 151 (1921), 79–111.

A. Sobczyk

1. Projections in Minkowski and Banach spaces, *Duke Math. J.*, 8 (1941), 78–106.

W. Sierpinski

1. *Leçons sur les nombres transfinis*, Gauthier-Villars, Paris, 1950.

M. H. Stone

1. *Linear transformations in Hilbert space*, American Mathematical Society Colloquium Publications, vol. 15, New York, 1932.

A. E. Taylor

1. Analysis in complex Banach spaces, *Bull. Amer. Math. Soc.*, 49 (1943), 652–669.
2. Spectral theory of closed distributive operators, *Acta Math.*, 84 (1950), 189–224.
3. New proofs of some theorems of Hardy by Banach space methods, *Math. Mag.*, 23 (1950), 115–124.
4. Banach spaces of functions analytic in the unit circle, *Studia Math.*, 11 (1950), 145–170 and 12 (1951), 25–50.

A. E. Taylor and C. J. A. Halberg Jr.

1. General theorems about a bounded linear operator and its conjugate, *J. Reine Angew. Math.*, 198 (1957), 93–111.

G. O. Thorin

1. *Convexity theorems generalizing those of M. Riesz and Hadamard, with some applications*, University of Lund thesis, Uppsala, 1948.

E. C. Titchmarsh

1. *Theory of Functions*, 2nd ed., Oxford University Press, 1939.
2. *Eigenfunction expansions*, Oxford University Press, 1946.

S. S. Walters

1. The space H^p with $0 < p < 1$, *Proc. Amer. Math. Soc.*, 1 (1950), 800–805.
2. Remarks on the space H^p, *Pacific J. Math.*, 1 (1951), 455–471.

J. V. Wehausen

1. Transformations in topological linear spaces, *Duke Math. J.*, 4 (1938), 157–169.

A. Weil

1. *Sur les espaces à structure uniforme*, Actualités Sci. Ind., No. 551, Hermann, Paris, 1937.

H. Weyl

1. The method of orthogonal projection in potential theory, *Duke Math. J.*, 7 (1940), 411–444.

D. V. Widder

1. *The Laplace Transform,* Princeton University Press, 1946.

N. Wiener

1. *The Fourier Integral,* Cambridge University Press, 1933.

A. Wintner

1. *Spektraltheorie der unendlichen Matrizen,* Hirzel Verlag, Leipzig, 1929.
2. Zur Theorie der beschränkten Bilinearformen, *Math. Zeit.,* 30 (1929), 228–282.

K. Yosida and E. Hewitt

1. Finitely additive measures, *Trans. Amer. Math. Soc.,* 72 (1952), 46–66.

A. C. Zaanen

1. *Linear analysis,* North Holland, Amsterdam, 1953.

A. Zygmund

1. *Trigonometrical Series,* Monografje Matematyczne, Warsaw-Lwów, 1935.

LIST OF SPECIAL SYMBOLS

$\in$	Is an element of, 1
$\notin$	Is not an element of, 1
$\emptyset$	The empty set, 1
$\subset$	Is a subset of, 1
$\{x:P(x)\}$	Set notation, 2
$\cup$	Union, 2
$\cap$	Intersection, 3
$C(E)$ or E'	Complement of E, 3
$E - F$	Set difference, 3
$X \times Y$	Cartesian product, 4
F^{-1}	Inverse function, 5
sup	Least upper bound, 5
inf	Greatest lower bound, 5
δ_{ij}	The Kronecker delta, 6
$C[a, b]$	Space of continuous functions, 14, 90
R_n	Real arithmetic space, 14
C_n	Complex arithmetic space, 14
$\mathscr{L}^p$	A class of functions, 16, 90, 376
L^p	A "Lebesgue" space, 17, 90, 376
$x = {}^0 y$	Equality almost everywhere, 17

$[x]$	Equivalence class notation, 17, 104
$BV[a, b]$	Space of functions of bounded variation, 17, 91
$\mathscr{D}(A)$	Domain of A, 18
$\mathscr{R}(A)$	Range of A, 18
$\langle x, x' \rangle$	Linear functional notation, 34
X^f	Algebraic conjugate of X, 34
X^{ff}	Algebraic conjugate of X^f, 36
$\dim X$	Dimension of X, 46
A^T	Transpose of A, 47
S^0	Annihilator, 48, 224
${}^0S$	Annihilator, 49, 225
J	Canonical mapping operator, 36, 191
$\mathscr{N}(A)$	Null manifold of A, 52
$\bar{S}$	Closure of S, 57
$\operatorname{int}(S)$	Interior of S, 58
$f^{-1}(T)$	Inverse image of T, 61
T_1 — space	A type of topological space, 59
T_1, T_2, T_3 — spaces	Types of topological spaces, 65
$d(x_1, x_2)$	Distance, 68
$\hat{X}$	Completion of X, 76, 99
Π	Cartesian product, 79
$\lVert T \rVert$	Norm of T, 86
$l^p(n)$, $l^\infty(n)$	Minkowski spaces, 88
l^p, l^∞	Sequence spaces, 88
(c)	A sequence space, 88, 201
$B(T)$ and $B[a, b]$	Function spaces, 89
$C(T)$	A space of continuous functions, 89
$\sup^0$	Essential supremum, 90
$\mathscr{L}^\infty$	A class of functions, 90
L^∞	A space formed from $\mathscr{L}^\infty$, 91
$V(x)$	Total variation of x, 91
$\mathfrak{M}_p[f; r]$	A mean value, 92
$\mathfrak{M}_\infty[f; r]$	A maximum value, 92
$\mathfrak{A}$	A class of analytic functions, 92
H^p, H^∞	Spaces of analytic functions, 93
CA	A space of analytic functions, 93
(x_1, x_2)	Inner product, 106
$\mathscr{T}(X, M)$	A weak topology, 152
(s)	A sequence space, 156
S	A function space, 156

$[X, Y]$	A space of operators, 162
sgn	Signum function, 193
$p' = \frac{p}{p-1}$	Conjugate index, 193
v^*	Normalization of v, 199
(c_0)	A sequence space, 201
$\nu(M)$	Characteristic of M, 202
$M_1 \oplus M_2$	Direct sum, 240
$\rho(T)$	Resolvent set, 253
$\sigma(T)$	Spectrum, 253
R_λ	Resolvent operator, 257
$r_\sigma(T)$	Spectral radius, 262
$C\sigma(T)$	Continuous spectrum, 264
$R\sigma(T)$	Residual spectrum, 264
$P\sigma(T)$	Point spectrum, 265
$\alpha(T)$	Ascent of T, 271
$\delta(T)$	Descent of T, 272
$\mathfrak{A}(T)$	A class of functions, 288
$\mathfrak{A}_\infty(T)$	A class of functions, 292
$\sigma_e(T)$	Extended spectrum, 298
$C_\infty(T)$	A space of continuous functions, 366, 387
$(T, \boldsymbol{S}, \mu)$	A measure space, 367
χ_E	Characteristic function, 368
$\mathscr{L}(\mu)$	A class of functions, 372
$L(\mu)$	A space formed from $\mathscr{L}(\mu)$, 372
$\int x\, d\mu$	Integral of x, 369, 372
$\lvert\mu\rvert$	Total variation measure, 374
$\mathscr{L}^p(\mu)$	A class of functions, 376
$L^p(\mu)$	A space formed from $\mathscr{L}^p(\mu)$, 376
$N(x)$	The support of x, 386
$S\int x\, d\mu$	S-integral of x, 402
$B^N(T)$	A space of functions, 404
$L\int x\, d\mu$	L-integral of x, 405

Discussions Relating to Particular Spaces

$B[a, b]$, 196
$B(T)$, 89, 102, 403
$BV[a, b]$, 17, 19, 91, 103, 198–200
(c), 51, 201, 217–221
(c_0), 201, 217–220

$C[a, b]$, 14, 19, 20, 27, 28, 35, 94, 96, 165, 172, 181, 195, 196, 200, 210, 276, 285, 291, 301
CA, 93, 94, 103
$C(T)$, 89, 94, 103
$C_\infty(T)$, 387, 393–400
$H^p(0 < p < 1)$, 92, 158, 212
$H^p(1 \leqslant p)$, 92, 93, 94, 103, 185
H^∞, 92, 93
l^2, 16, 35, 108, 110, 115, 120, 121, 286, 327
l^p, 88, 100, 102, 182, 183, 194, 201, 210, 219, 220, 224, 265, 266, 278, 286, 300
l^∞, 88, 201, 217, 219, 220
L^2, 33, 108, 110, 112, 121, 122, 169, 170, 176, 277, 284, 301, 325, 326, 363
$L^p\ (0 < p < 1)$, 128, 143, 158, 210
$L^p\ (1 \leqslant p)$, 16, 29, 31, 32, 36, 90, 103, 185, 201, 377, 382 (*see also* $L(\mu)$ and $L^p(\mu)$)
L^∞, 91, 104, 381
$L(\mu)$, 366–373
$L^p(\mu)$, 376–385
(s), 156, 158
S, 156

INDEX

Absolutely continuous measure, 376
Absorbing set, 124
Absorption of a set, 128
Accumulation point, 58, 69, 70, 71
Achieser, N. I., and I. M. Glasmann, 122, 158, 252, 344, 364, 407
Adjoint operator, 160, 214, 249
Adjoint space, 186
Algebra, 162
Algebraic conjugate, 34
Algebraic reflexivity, 37, 45, 53, 192
Algebraic saturation, 51, 53
Almost everywhere, 17
Almost-periodic functions, 342
Analytic functions, spaces of, 31, 92, 93, 103, 150, 158
  with vector values, 205, 206, 207, 211, 213, 261, 314
Annihilator, 48, 49, 224, 225, 226
Approximation by polynomials, 94
Arithmetic space, 14
Ascent, 271, 279, 282
Ascoli's theorem, 276
Axiom of choice, 39
Bade, W. G., *296*, *320*, *407*
Baire's category theorem, 76, 98
Balanced hull, 124, 132
Balanced set, 123
Banach, S., 8, 55, 80, 157, 158, 201, 230, 232, 252, 287, 406, 407
Banach algebra, 366
Banach space, 98
Base, at a point, 58
  at 0, 124, 125, 146, 147
  for a space, 58, 60
Basis, countable, 207, 219, 220
  finite, 12, 13, 37, 38
  (*see also* Hamel basis)
Bessel's inequality, 106, 109, 116
Bilinear form, 322
Birkhoff, G., 406, 407
Boas, R. P., 296, 407
Borel measurability, 386
Borel measure, 386
  regular, 386, 389
  signed or complex, 388
Borel set, 366, 368
Boundary of a set, 58, 60

Boundary-value problems, 27
Bounded linear operator, 163
Bounded set, 70, 83, 127, 128, 151, 154, 157, 208
Bound of an operator, 86
Bourbaki, N., 55, 80, 123, 129, 158, 159, 406, 407

Canonical imbedding, 36
Canonical mapping, 36, 104, 191
Cantor's intersection theorem, 76
Cardinal number, 46, 116, 117, 118, 121
Cartesian product, 4, 79, 80
Category, 63, 76, 178, 179
Cauchy domain, 288
Cauchy sequence, 74
Cauchy's inequality, 6
Cauchy's integral formula, 32, 33, 211, 212, 261, 290
Cauchy's theorem, 211, 261, 289
Cayley transform, 362
Characteristic function, 368
Characteristic of a manifold, 202, 208
Characteristic value, 284
Charge, 401
Churchill, R. V., 122, 407
Clarkson, J. A., 231, 408
Closed-graph theorem, 161, 181
Closed operator, 160, 175
Closed set, 57
Closure of a set, 57, 60
Compactification, 67, 68
Compactness, 62, 63, 67, 70, 71, 72, 80
  and boundedness, 70
  and closedness, 62, 66, 70
  and convexity, 134, 151
  and separability, 71
  conditional, 63
  conditional sequential, 286
  local, 67, 68
  relative, 63
  weak, 227–230, 232
Compact operator, 274, 335, 342
  and weak convergence, 287, 342
  conjugate of, 275, 282, 286
  examples of, 276, 277, 278, 286, 342
Compact resolvent, 342–345
Complement, 3
Complementary manifold, 241
Completely continuous operator, 274
Completeness, 74, 98, 129
  proofs of, 100–103
  (*see also* Weak sequential completeness)
Complete orthonormal set, 114, 115, 117, 121–122
Complete regularity, 66
Completion of a space, 75, 98, 99
Complex linear space, 9
Complex measure, 375
Components of a vector, 24
Congruence (of spaces), 84, 85
Conjugate, algebraic, 34
  normed, 185
  topological, 54
Conjugate index, 193
Conjugate operator, 160, 213, 214
Conjugate space, 160, 185
Conjugation, 118
Content, 387
Continuity, and closed sets, 62
  and compactness, 63
  and open sets, 61
  uniform, 70
Continuous function, 61, 70
Continuous spectrum, 264, 353
Convergence in measure, 370
Convergent sequence, 69
Convex body, 142, 143
Convex hull, 131, 132, 133, 134
Convex function, 221, 223
Convex sets, 130–136, 140–143, 151, 154, 230, 243, 330
Cooke, R. G., 158, 252, 364, 408
Countability, axioms of, 64
Countable additivity, 374
Countable base, 64, 69
Courant, R., and D. Hilbert, 122, 320, 341, 408
Covering, 62

Daniell integration, 385
Day, M. M., 143, 408
Deficiency index, 362
Degenerate kernel, 319
Denseness, 63
Descent, 272, 279
Diameter of a set, 76
Dieudonné, J., 157, 159, 232, 408

Difference (set-theoretic), 3
Differential equations, 15, 19, 20, 27, 28, 30, 123, 171–175, 176, 181, 286, 291, 295, 321, 326, 344, 345
Dimension, 11, 12, 13, 46, 191
Directed sets, 129, 401
Direct sum, 240
Dirichlet problem, 26, 161, 188, 247, 248, 285
Dirichlet's principle, 247
Discrete topology, 57, 60
Disjoint sets, 64
Distance between sets, 72
Distance function, 68
Dixmier, J., 114, 408
Domain of a function, 4
Dominated convergence, 373
Dual basis, 38
Duality, 153
Dual space, 186
Dunford, N., 158, 320, 355, 356, 364, 406, 408
  and B. J. Pettis, 313, 408
  and J. T. Schwartz, 222, 231, 252, 320, 406, 408

Eberlein, W. F., 230, 408
Eigenmanifold, 255
Eigenvalue, 255, 265, 284
Eigenvector, 255
Empty set, 1
Equicontinuity, 276
Equilibré, 123
Essential lub, 90
Essential singularity, 306
Euclidean space, 118, 119
Existence problem, 19, 20, 26
Extended spectrum, 298
Extension, 40
  closed linear, 259, 260
  minimal, 260
Extension theorems, 40–44, 144, 148

Family, 2
Fatou's lemma, 377
Fejér, L., 94
Fibonacci numbers, 268
Fichtenholz, G., and L. Kantorovitch, 406, 408
Filters, 129
Finite-dimensional spaces, 11, 12, 13, 21, 22, 23, 37–39, 45, 94–98, 127, 129, 316, 318, 319
Fourier coefficients, 32, 112, 121, 296
Fourier-Plancherel transform, 360, 363
Fourier series, 33, 94, 109, 121
Fourier transform, 8
Fréchet, M., 80, 157, 158, 408
Fredholm, I., 27, 168, 285
Friedman, B., 252, 320, 364, 408
Function, 4
  continuous, 61
  domain of, 4
  essentially bounded, 90
  inverse, 5
  N-bounded, 404
  of bounded variation, 17, 35, 91, 195, 200
  range of, 4
Functional notation, 4

Gram-Schmidt process, 106, 116, 117
Graph of a function, 175
Graves, L. M., 276, 408
Greatest lower bound, 5
Green's function, 27, 188, 344
Grimshaw, M. E., 55, 319, 409
Group, 10

Hadamard's three-circles theorem, 211
Hahn-Banach theorem, 8, 82, 161, 186, *187*, *188*
Halberg, C. J. A., Jr., 238, 268, 408, 411
Half space, 140
Hall, D. W., and G. L. Spencer, 80, 408
Halmos, P. R., 55, 157, 158, 252, 320, 364, 366, 371, 373, 374, 376, 377, 378, 381, 385, 387, 406, 409
Hamburger, H. L., and M. E. Grimshaw, 55, 319, 409
Hamel, G., 44, 409
Hamel basis, 44, 45, 46, 150, 157, 240
Hamilton-Cayley theorem, 318
Hardy, G. H., 212, 409
Hardy, G. H., J. E. Littlewood, and G. Pólya, 5, 158, 212, 222, 409
Harmonic functions, 26
Hausdorff, F., 20, 330

Hausdorff space, 60, 64, 65, 66, 67, 126
Hellinger, E. D., 268, 364, 409
Hellinger, E., and O. Toeplitz, 164, 252, 341, 364, 409
Helly, E., 231
Hermite functions, 122
Hermite polynomials, 122
Hermitian form, 323
Hermitian operator, 321, 324
Hewitt, E., 406, 412
Hilbert, D., 16, 83
  (*see also* Courant and Hilbert)
Hilbert-Schmidt theory, 321
Hilbert space, 16, 55, 83, 118, 120, 243–245, 329–333, 345–364
Hildebrandt, T. H., 406, 409
Hille, E., 252, 296, 320, 409
Hölder's inequality, 6, 377
  converse of, 366, 379–380
Homeomorphism, 62
  linear, 85, 126
Homomorphism, 290
Hyers, D. H., 159, 409
Hyperplane, 137, 138–142

Inequalities, 6
Inner product, 82, 106, 108
Initial-value problem, 172
Integral equations, 27, 123, 165–175, 255, 284–285, 321, 341
Integrable function, 367, 369, 372
Integral (general definition of), 368, 369, 372
Integro-differential equation, 175
Interior-mapping principle, 180
Interior of a set, 58
Intersection of sets, 3
Invariant metric, 154, 155, 184
Invariant subspace, 268
Inverse function, 5
Inverse operator, 18, 86
Isometric isomorphism, 84
Isometric operator, 361
Isometric spaces, 74
Isomorphism, of linear spaces, 23, 24
  topological, 85, 86, 126

Jacobson, N., 46, 158, 409
Jensen's inequality, 6
Jordan decomposition, 374, 389, 391
Jordan normal form, 319
Julia, G., 158, 252, 409

Kakutani, S., 231, 409
Kantorovitch, L., 406, 408
Kelley, J. L., 55, 67, 80, 129, 157, 401, 409
Kellogg, O. D., 285, 409
Kernel, 165
  iterated, 167
  $\mathscr{L}^2$, 169, 277
  reciprocal, 167
  resolvent, 167
  Volterra, 168
Kolmogoroff, A. N., 82
Kronecker delta, 6

Laguerre functions, 122
Laguerre polynomials, 122
Laplace's equation, 26
Laplace transform, 8
Lattice, 391
Laurent theorem (series expansion), 211, 261, 262, 305, 313
Lax, P. D., 190, 409
Least upper bound, 5
Lebesgue's dominated convergence theorem, 373
Legendre polynomials, 25, 122
Limit of a sequence, 69
Linear combination, 11
Linear dependence, 10, 11
Linear functional, 33
Linear homeomorphism, 85, 126
Linear independence, 11, 240
Linear manifold, 10, 84
Linear operator, 17, 18
Linear problem, 7, 8, 19
Linear space, 8, 9
Linear variety, 137, 139
Liouville, J., 164
Liouville's theorem, 211, 261, 264, 314
Livingston, A. E., 158, 409
Ljusternik, L. A., and W. I. Sobolew, 80, 158, 252, 320, 364, 410
Local analyticity, 288
Local convexity, 145, 146
Locally bounded space, 157, 158
Loomis, L. H., 67, 80, 252, 385, 406, 410
Lorch, E. R., 320, 410
Löwig, H., 46, 410

Mackey, G. W., 159, 410
McShane, E. J., 276, 410
Manifold, *see* Linear manifold
Matrices, 22, 23
  infinite, 183, 217–221
Maximal element, 40
Maximal subspace, 137
Maximum-modulus theorem, 211
Measure, convergence in, 370
Measurable function, 368, 402
Measure on a ring, 367
Measure space, 367
Metric linear space, 155, 179–181
Metric space, 68
  of measurable sets, 381
Metrizability, 150, 157, 191
Milman, D., 231, 410
Mimura, Y., 231
Minimal polynomial, 316
Minkowski's inequality, 6, 278, 377
Minkowski functionals, 134–135
Miranda, C., 190, 410
Moment problem, 20
Moment sequence, 19
Montel, P., 150
Moore-Smith convergence, 129, 401
Murray, F. J., 242, 410

Nagy, B. v. Sz., 158, 364, 410
  (*see also* Riesz and Nagy)
Neighborhood, 58
Nets, 129, 401
Neumann, C., 164
von Neumann, J., 158, 159, 364, 410
Neumann expansion, 164, 169
Neumann problem, 285
Nilpotent operator, 318, 320
Nonseparable spaces, 89, 91, 92, 93, 120, 188
Norm, of a functional, 185
  of an operator, 86, 162
  of a vector, 82, 83
Normable space, 82, 136, 157, 158
Normalized function of bounded variation, 199, 200
Normal operator, 251, 329, 331, 332
Normal space, 65, 67, 73
Norm-determining manifold, 202, 204, 208
Normed conjugate, 185
Normed vector space, 83

Null manifold, 52, 271

Open set, 57, 60
Operational calculus, 290, 296, 352
Operator, 17
  of Fredholm type, 168
  of Volterra type, 168
Orthogonal complement, 161, 245, 246
Orthogonality, 109, 242
Orthogonal projection, 245, 246, 248, 333
Orthonormal set, 109, 110–118
Outer measure, 367, 388

Parallelism of hyperplanes, 137
Parseval's identity, 106, 109, 116, 121
Partially ordered linear space, 391
Partial ordering, 39
Partition, 401
Perfect separability, 64, 69
Pettis, B. J., 231, 410
  (*see also* Dunford and Pettis)
Phillips, R. S., 313, 410
Point spectrum, 265, 353
Pole of resolvent, 306, 310, 311, 312, 313
Pollard, H., 356, 410
Positive linear functional, 387
Precompactness, 78, 134
Pre-Hilbert space, 119
Product of sets, 4, 79
Product space, 79, 80
Projection, 240, 241, 245, 333
Projection theorem, 161
Proper subset, 1
Proper subspace, 10

Quadratic form, 322
Quotient space, 104, 105, 130, 192, 227

Radon-Nikodym theorem, 376
Range of a function, 4
Real linear space, 9
Reducibility, 268–269, 299
Reflexivity, algebraic, 37, 45, 53, 192
  of normed linear spaces, 192, 208, 209, 210, 215, 227, 229, 230, 231, 232, 233, 240, 385
Regularly closed subspace, 225
Regular space, 64
Relatively bounded linear functional, 393
Relative topology, 60

Rellich, F., 342, 410
Representation of a space, 54
Residual spectrum, 264
Resolution of the identity, 322, 350, 354, 360
Resolvent of an operator, 254, 257, 262
Resolvent set, 253, 257
Riesz, F., 32, 93, 96, 104, 212, 222, 320, 410
Riesz, F., and B. v. Sz. Nagy, 201, 252, 285, 320, 361, 364, 406, 410
Riesz, M., 222, 411
Riesz convexity theorem, 221–222
Riesz-Fischer theorem, 104, 106, 112
Riesz's lemma, 96, 97, 98
Ring of sets, 367
  σ-ring, 367
Rosenbloom, P. C., 401, 405, 406, 411

Saks, S., 374, 411
Saturated subspace, 51, 53, 225, 226, 231–233
Scalars, 9
Schmeidler, W., 158, 364, 411
Schur, I., 327, 411
Schwartz, J. T., *see* Dunford and Schwartz
Schwarz's inequality, 6, 107, 277
Second conjugate space, 191
Self-adjoint operator, 251, 255, 329
  unbounded, 362
Seminorms, 134, 143–147, 149, 150
Separability, 64, 69
Separable spaces, 64, 69, 71, 88, 89, 90, 93, 94, 187, 191, 209, 381
Separation axioms, 59, 64
Separation theorems, 142, 148, 151, *186*
Sets, 1
  absorbing, 124
  bounded, 70, 83, 127, 128, 151, 154, 157
  closed, 57
  compact, 62
  completely ordered, 40
  complete orthonormal, 114, 115, 117, 121–122
  convex, 130–136
  open, 57, 60
  orthonormal, 109, 110–118
  partially ordered, 39
  precompact, 78, 134
Sierpinski, W., 118, 411
Signed measure, 374
Simple function, 368
Sobczyk, A., 242, 411
Space, 8
  complete, 74
  completely regular, 66
  Euclidean, 118, 119
  Hausdorff, 60, 64, 65, 66, 67, 126
  inner-product, 82, 106, 108
  linear, 8, 9
  locally bounded, 157, 158
  locally compact, 67, 68
  locally convex, 145, 146, 150
  metric, 68
  metric linear, 155
  normal, 65, 67
  normed linear, 82, 83
  regular, 64, 126
  second conjugate, 191
  T_1, *59, 64, 149, 153*
  T_1–T_4, 65
  topological linear, 81, 123–130
  uniformly convex, 231
  unitary, 119
Spectral-mapping theorem, 262, 263, 302 303
Spectral radius, 262, 263, 331
Spectral set, 298
Spectral theorem, 321, 332, 345, 356, 357
  for compact symmetric operators, 336–341
  for unbounded operators, 362
Spectral theory, 253
Spectrum, 253, 257, 261, 264
  continuous, 264
  extended, 298
  point, 265
  residual, 264
Spencer, G. L., 80, 408
Sphere, 68
State diagram, 161, 236, 237
State of an operator, 235, 236
State of a pair, 236
Stieltjes integral, 19, 35, 195, 197
Stone, M. H., 158, 242, 252, 320, 330, 344, 364, 411
Sturm-Liouville equation, 175, 321
Subalgebra, 162
Subbase, 59
Sublinear functional, 42

Subset, 1
Subspace, 10, 84
Summability (C, 1), 94
Support of a function, 386
Support of a set, 142
Symmetric bilinear form, 323
Symmetric operator, 324
Symmetric projection, 333
Symmetric set, 123

Taylor, A. E., 94, 213, 289, 297, 320, 411
Taylor, A. E., and C. J. A. Halberg, Jr., 238, 411
Taylor's theorem, 261, 262
Thorin, G. O., 222, 411
Titchmarsh, E. C., 211, 320, 344, 405, 411
Toeplitz, O., 330
  (*see also* Hellinger and Toeplitz)
Topological conjugate, 54
Topological linear space, 81, 123–130
Topological space, 56, 57
Topologies, 57, 59, 60, 70, 125, 128, 146, 147
Totally charged space, 401
Total set of functionals, 152, 153, 157
Total variation, 91, 196
Total variation measure, 374
Translation, 137
Transpose of a matrix, 48
Transpose of an operator, 47, 48
Triangular inequality, 68, 107
T_1-space, 59
T_2-space, 60
T_2–T_4 space, 65
Two-point problem, 173
Tychonoff's theorem, 80

Uniform boundedness, principle of, 201–204
Uniform continuity, 70
Uniform convexity, 231
Uniform topology, 163
Union, 2
Uniqueness problem, 19, 20, 26
Unit, 162
Unitary operator, 356
Unitary space, 119
Upper bound, 40
Urysohn's lemma, 67

Vector lattice, 391
Vector space, 9
Volterra, V., 27, 168

Walters, S. S., 158, 411
Weak compactness, 227–230, 232
Weak convergence, 208, 210, 230, 287
Weak sequential completeness, 210
Weak topology, 127, 149, 151–154, 161, 227–233
Weak*-topology, 209
Wehausen, J. V., 159, 411
Weierstrass theorem, 94, 346, 347
Weil, A., 126, 411
Weyl, H., 248, 249, 411
Widder, D. V., 20, 412
Wiener, N., 122, 412
Wintner, A., 252, 330, 364, 412

Yosida, K., and Hewitt, E., 406, 412

Zaanen, A. C., 55, 158, 252, 313, 320, 364, 406, 412
Zorn's lemma, 39, 40, 41, 44, 45, 55, 114
Zygmund, A., 32, 93, 94, 121, 222, 412